

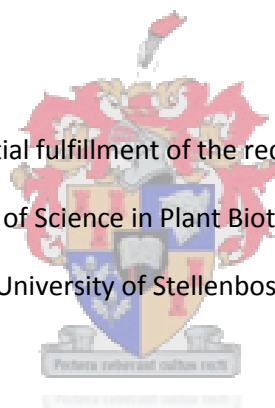
Identification of the Genes Encoding Enzymes Involved in the Synthesis of the Biopolymer Paramylon from *Euglena gracilis*

Stephen Mackay

Thesis presented in partial fulfillment of the requirements for the degree

Masters of Science in Plant Biotechnology

University of Stellenbosch



Supervisor: Prof. Jens Kossmann

Faculty of Science

Department of Genetics

December 2010

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

A handwritten signature in black ink, appearing to read 'S Mackay', written in a cursive style.

S Mackay

December 2010

Copyright © 2010 University of Stellenbosch

All rights reserved

Abstract

Recent advances in medical pharmacology have identified the immune-potentiating effects of β -1,3-glucans on mammalian immune systems. Extensive research has identified and described the mechanisms of action and receptor binding specificity of different β -1,3-glucans as well as their structural and functional relationships. Molecular mass, solubility, structural order, degree of branching as well as chemical modification all determine the effectiveness of the β -1,3-glucan immune-modulating activities, which typically include; macrophage activation, antibody adjuvant activities, reduction of LDL-cholesterol, leukocyte mitogenic activities, cytokine and chemokine production as well as antiviral and antitumor activities. Currently β -1,3-glucans have been sold commercially under the name β -glucan, mostly in the form of Betafectin, a genetically modified yeast derived β -1,3-glucan.

Recent studies of different β -1,3-glucans have identified the pharmacological activities of paramylon, a *Euglena* derived β -1,3-glucan. Although paramylon has relatively low immune-stimulating activities, chemical modification of the paramylon granule increased immune-potential with specific antimicrobial and anti-HIV activities. Due to these specific immune-potentiating activities, paramylon is novel in terms of both structure as well as functional activity.

In terms of biotechnological application, paramylon is greatly favoured as it is synthesized as an insoluble membrane bound granule in the cytosol of *Euglena* where most plant and fungal β -1,3-glucan synthases are cell membrane bound highly regulated multifunctional complexes, synthesizing β -1,3-glucan as cell wall components. Due to the novel granular nature of paramylon, expression in other systems with genetic modification could potentially further increase immuno-potentiating activities.

In this study, different approaches were attempted in order to identify the genes involved in paramylon synthesis including; constructing and screening a *Euglena gracilis* cDNA library, sequence analysis of the purified proteins as well as transcription analysis of the sequenced transcriptome and genome of *E. gracilis*. Putative candidates that encode subunits of the paramylon synthase complex have been identified.

Opsomming

Onlangse vordering in mediese farmakologie het die immuun-stimulerende effek van β -1,3-glukane op die soogdier immuunsisteem geïdentifiseer. Intensiewe navorsing het die meganisme van die werking en reseptor bindingspesifisiteit van verskillende β -1,3-glukane, asook hulle struktuur en funksionele verhoudings, geïdentifiseer. Die molekulêre massa, oplosbaarheid, strukturele orientasie, mate van vertakking asook chemiese modifikasies bepaal almal die effektiwiteit van die β -1,3-glukaan immuun-modulerende aktiwiteite. Tipiese immuno-moduleringsaktiwiteite sluit makrofaag aktivering, teenliggaampie adjuvant aktiwiteite, verlaging van LDL-cholesterol, leukosiet mitogeniese aktiwiteite, sitokien en chemokien produksie asook anti-virale en antitumor aktiwiteite in. Huidiglik word β -1,3-glukane onder die naam β -glukaan verkoop meestal in die vorm van Betafectin, 'n geneties gemodifiseerde gis wat van β -1,3-glukaan afkomstig is.

Onlangse studies van verskillende β -1,3-glukane het die farmakologiese aktiwiteit van paramylon, 'n *Euglena* afkomstige β -1,3-glukaan geïdentifiseer. Alhoewel paramylon relatiewe lae immuun-stimulerende aktiwiteit toon, verhoog chemiese modifikasies van die paramylon granules immuun-stimulering, spesifiek die anti-mikrobiese en anti-MIV aktiwiteite. Weens hierdie spesifieke immuun-stimulerende aktiwiteit, word paramylon as nuut beskou veral in terme van beide struktuur asook funksionele aktiwiteit.

In terme van biotegnologiese toepassing, verkry paramylon voorkeur aangesien dit as 'n onoplosbare membraangebonde granule in die sitosol van *Euglena* gesintetiseer word terwyl meeste plant en fungus β -1,3-glukaan sintases hoogs gereguleerde multifunksionele selmembraan gebonde komplekse is wat β -1,3-glukaan asook ander selwand komponente sintetiseer. Weens die unieke granulêre natuur van paramylon, sal uitdrukking in ander sisteme 'n moontlike industrie skep waar deur die transgeniese uitdrukking van granulêre paramylon verdere verbetering van die immuun-stimulerings aktiwiteite kan lei.

In hierdie studie is verskillende benaderings aangewend om die gene wat by paramylon sintese betrokke is te identifiseer, dit sluit in die konstruksie en sifting van 'n *E. gracilis* cDNS biblioteek, aminosuur volgorde analise van gesuiwerde proteiene asook die transkripsionele analise van die volgorde van die transkriptoom en genoom van *E. gracilis*. Moontelike kandidate wat vir die subeenhede van die paramylon syntase kompleks kodeer is geïdentifiseer.

Acknowledgements

I would like to thank:

Prof. Jens Kossman for supervising this project, for proofreading and guiding this project and giving me the opportunity to study.

Dr. Jan Bekker, for all his assistance, ideas, patience and time.

Staff, students and friends at the IPB.

My family and friends for their continual patience, love and support.

Table of Contents

Abstract	iii
Opsomming	iv
Acknowledgements	v
List of Figures	ix
List of Tables	x
Abbreviations	xi
Review of Literature	1
1.1 <i>Euglena gracilis</i>	1
1.2 Taxonomy	2
1.3 Paramylon and paramylon synthesis	3
1.4 Glycosyl transferases	6
1.5 Partial purification of β -1,3-glucan synthases and identification of catalytic subunits	8
1.6 Paramylon synthase	11
1.7 Cytokine-related immunopotentiating activities of paramylon	11
1.8 Aims and experimental approach	13
Materials	15
2.1 Chemicals and kits	15
2.2 Primers	15
2.3 Primers designed from an isolated <i>E. gracilis</i> β -glucan synthase sequence	17
2.4 Web – based programs	17
Methods	18
3.1.1 Construction and screening of the cDNA library	18
3.1.1.1 Growth conditions of <i>Euglena gracilis</i>	18
3.1.1.2 Total RNA extraction	18
3.1.1.3 Isolation and quantification of mRNA	18
3.1.1.4 cDNA synthesis	19
3.1.1.5 cDNA library construction and in vitro packaging of the phage library	19
3.1.1.6 Amplification of the cDNA primary library	19
3.1.1.7 Mass excision of the amplified library	19
3.1.1.8 Single clone excision	20
3.1.2 Screening of the phage library	20

3.1.2.1	Infection and Expression of the phage library.....	20
3.1.2.2	Screening of the phage library using an antibody.....	20
3.1.2.3	Screening of the phage library using radio-labeled substrate	21
3.1.2.4	Functional screening of the mass excision library using aniline blue.....	22
3.1.2.5	Screening of the mass excision library using cell lysis.....	22
3.1.3	cDNA synthesis for PCR screening and molecular cloning	23
3.1.3.1	Design of PCR primers	23
3.1.3.2	PCR conditions.....	23
3.1.3.3	DNA purification, cloning, transformation and sequencing.....	24
3.2	Partial purification of paramylon synthase proteins.....	24
3.2.1.1	Protein extraction.....	24
3.2.1.2	Dialysis of the crude protein extract	25
3.2.1.3	Protein quantification.....	25
3.2.2	FPLC, PAGE separation, product entrapment and immunoblotting of partially purified paramylon synthase proteins	25
3.2.2.1	FPLC of paramylon synthase proteins	25
3.2.2.2	Detection of in gel paramylon synthase activity	26
3.2.2.3	Partial purification of paramylon synthase through product entrapment	26
3.2.2.4	MOPS PAGE of the product entrapped protein	27
3.2.2.5	Staining of PAGE gels.....	27
3.2.2.6	Western blots	28
3.2.3	Tryptic digestion and LCMSMS of partially purified paramylon synthase	28
3.2.3.1	Trypsin digestion of excised gel fragments	28
3.2.3.2	LCMSMS.....	29
3.3	Transcriptome sequencing	29
Results and Discussion.....		29
4.1.1	Construction, amplification, infection and expression of the cDNA library.....	29
4.1.2	Screening of the library	30
4.1.2.1	Screening of the phage library using an antibody.....	30
4.1.2.2	Screening of the library using radio-labeled substrate	31
4.1.2.3	Functional screening of the mass excision library using aniline blue.....	32
4.1.2.4	Colony library screens	33
4.1.3	Screening of cDNA using PCR	33
4.2	Protein purification.....	36

4.2.1	Paramylon synthase crude protein extraction	36
4.2.2	FPLC, PAGE separation, product entrapment and immunoblotting of partially purified paramylon synthase proteins	36
4.2.2.1	FPLC of paramylon synthase proteins	36
4.2.2.2	Detection of in gel paramylon synthase activity	37
4.2.2.3	Paramylon synthase partial purification through product entrapment.....	38
4.2.2.4	MOPS-PAGE of the product entrapped protein	39
4.2.2.5	Western Blot and gel excision of paramylon synthase proteins	39
4.2.3	LCMSMS.....	40
4.3	Sequence analysis of the <i>E. gracilis</i> transcriptome and genome	41
4.3.1	Sequencing of the <i>Euglena gracilis</i> transcriptome.....	41
4.3.2	BLASTing the partially completed <i>Euglena gracilis</i> genome sequence	44
4.3.2.1	Contig 11792.....	49
4.3.2.2	Contig 11487.....	50
General discussion, conclusion and future work		52
Reference list		54
Addendum A: Predicted peptide LCMSMS sequences of the purified paramylon synthase proteins.		70

List of Figures

Figure 1: <i>Euglena gracilis</i>	1
Figure 2: The phylogeny of <i>E. gracilis</i>	3
Figure 3: The structure of the paramylon granule.	4
Figure 4: FPLC purification of polyclonal anti-paramylon synthase antibodies.	37
Figure 5: FPLC purified paramylon synthase proteins.....	37
Figure 6: Schiff stained Native PAGE of in-gel paramylon synthase activity	38
Figure 7: Western Blot and MOPS PAGE analysis of partially purified product entrapt proteins	40
Figure 8: Purified double stranded cDNA smears sequenced using the genome sequencer FLX system.	41
Figure 9: Read length distribution of the sequenced transcriptome.....	43
Figure 10: Contig length distribution of the sequenced transcriptome.....	43
Figure 11: Contig queries from the sequenced genome with partial identity to glycosyl transferases.	48
Figure 12: CESA_CelA like cellulose synthase GT-2 sequence from <i>Verticillium albo-atrum</i>	49
Figure 13: Contig 11792..	49
Figure 14: Contig 11487 ORF +3.....	51
Figure 15: Contig 11487 ORF +2.....	51

List of Tables

Table 1: Methods of purification and identification of β -1,3-glucan synthase proteins in literature.	10
Table 2: Conserved β -1,3-glucan synthase sequences used to design primers.	16
Table 3: Degenerate primers designed from β -1,3-glucan synthase conserved domains.	16
Table 4: Primers based on th <i>D. hanseii</i> glucan synthase peptide.	17
Table 5: Website-based bioinformatics tools.	17
Table 6: BLASTx results of false positive clones of the [14 C] UDP-glucose radio-labeled phage screens.	32
Table 7: PCR results of cDNA and the mass excision library amplified and sequenced.	35
Table 8: Contig assembly of the transcriptome	42
Table 9: Predicted peptide sequences aligned with the contigs of the sequenced genome	46
Table 10: BLASTx results of the genome contigs aligned with two predicted peptide sequences.	46
Table 11: Contig sequences from the sequenced genome with partial identity to GT-2 glycosyl transferase protein sequences.	47

Abbreviations

⁰ C	degrees centigrade
¹⁴ C	Carbon-14 radio-isotope
bp	base pairs
BSA	Bovine serum albumin
cDNA	complementary deoxyribonucleic acid
CHAPS	3-[(3-cholamidopropyl)-dimethylammino]-1-propane sulfonate
CNBr	cyanogen bromide
CTAB	cetyltrimethylammonium bromide
DMSO	dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DNAse I	Deoxyribonuclease I (EC 3.1.21.1)
DTT	dithiotreitol
E- value	Expect value
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tag
FPLC	flow pressure liquid chromatography
g	gram
<i>g</i>	gravitational force
GT	glycosyl transferase
GTP	guanidine triphosphate
h	hour
IPTG	isopropyl- β-D-thiogalactopyranoside
kDa	kilo Dalton
L	litre
LB	Luria broth

LCMSMS	liquid chromatography mass spectrophotometry
mg	milligram
min	minute
ml	milliliter
μl	microlitre
μM	micromolar
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger ribonucleic acid
m/v	mass per volume
nm	nanometers
OD ₆₀₀	optical density at 600 nm
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
RT	room temperature
s	second
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
U	enzyme units
UDP	uridine diphosphate
v/v	volume per volume
w/v	weight per volume

Review of Literature

1.1 *Euglena gracilis*

Euglenoids (Euglenophyceae, Euglenophyta), consist of around 40 genera and 1000 species (van den Hoek *et al.*, 1995). *Euglena*, the most common fresh water genus, are autotrophs, although some exist as obligate heterotrophs. They grow normally as spindle-shaped cylindrical cells and can vary in length between 10 μm and 500 μm . Euglenids have been difficult to place phylogenetically due to their unusual morphology. The photoautotrophic euglenoids and the secondary heterotrophs are characterized by their specific pellicle, paramylon, a β -1,3-glucan (Barras and Stone, 1968), as well as the euglenoid specific closed mitosis. In phototrophs, *Euglenae* are characterized by chloroplasts enveloped by three membranes which contain chlorophylls *a* and *b*, β -carotene, antheroxanthin, neoxanthin and catanoid and quinones (Leedale, 1967), a phototactic photoreceptor (Dawson and Walne 1991b) and a separate extraplastidial eyespot (Kreimer, 1994; Leedale, 1967; Walne and Arnott, 1967; Walne and Kivic, 1990).

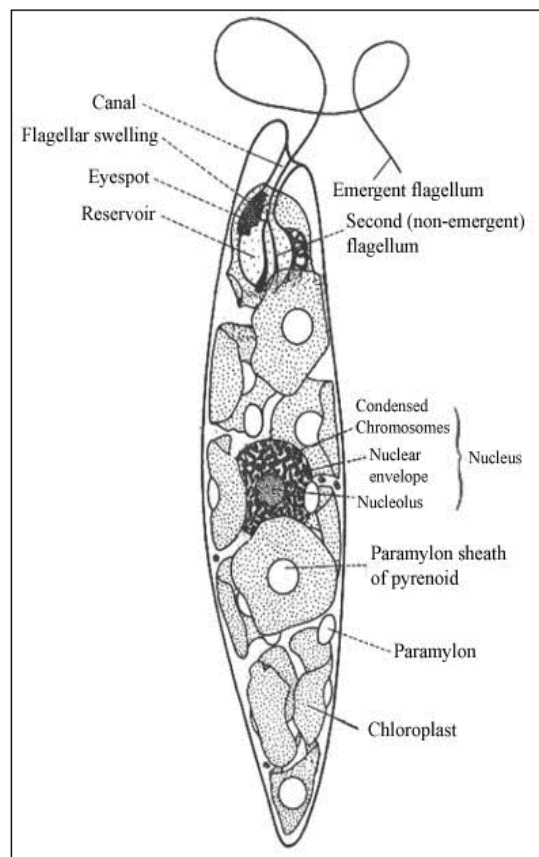


Figure 1: *Euglena gracilis*. (www.botany.ubc.ca/Biol320/ultra/whol03.htm)

Besides paramylon synthesis, *E. gracilis* (Figure 1) is one of the few microorganisms which synthesize high amounts of antioxidant vitamins such as β -carotene, L-ascorbic acid and α -tocopherol simultaneously (Takeyama *et al.*, 1997) and produce relatively large amounts of wax esters (Inui *et al.*, 1983) and highly unsaturated fatty acids (Hayashi *et al.*, 1993) which is valued in the food and feed industries (Kitaoka and Hosotani, 1977). *Euglena* naturally rely on decaying organic matter as food source and grow on media consisting of acetate, glucose, glutamate, succinate, pyruvate and even high concentrations of ethanol in light and dark conditions as well as ammonia (Kusmic *et al.*, 1999) and are able to grow under harsh acid conditions as low as pH 1.8 (Yamane *et al.*, 2001).

1.2 Taxonomy

Euglenophyceae (Euglenozoa) are part of a diverse group of flagellate protists consisting of *euglenoids*, *kinetoplastids*, *diplonemids* and *postgaardi* all identified and characterized by the unique root pattern and paraxonemal cylinders in the flagella (Fenchel *et al.* 1995; Patterson, 1999; Simpson, 1997; Simpson *et al.* 1997). Cavalier-Smith (2003a, 2003b) placed the phylum Euglenozoa in the kingdom of Protozoa, subkingdom Bicilata and intrakingdom Excavata. Photosynthetic euglenoids and secondary hetrotrophs form the Euglenophyceae clade with two orders: Euglenales and Eutreptiales. Marine Eutreptiales (Eutreptia and Eutreptiella) consist of biflagellate and quadriflagellate taxa. Fresh water Euglenales form nine clades and two individual branches. *Euglena* are polyphyletic and split into four independent clades and two individual branches.

Most photoautotrophic eukaryotes inherited photosynthetic abilities from a cyanobacterial endosymbiont (McFadden, 2001), from which the genome of the endosymbiont was reduced to a plasmid within the chloroplasts of red and green algae, glaucocystophytes and higher plants (Adl *et al.*, 2005). While several other independent eukaryotic lineages adopted photosynthetic capabilities from another secondary endosymbiotic event where a eukaryotic algae was engulfed by the host eukaryote (Gibbs, 1978; Stoebe and Maier, 2002). Where both endosymbiotic events resulted in endosymbiotic gene transfer between host chromosomes and organelles (Archibald *et al.*, 2003; Martin *et al.*, 1993; Timmis *et al.*, 2004). Endosymbiotic gene transfer can vary in frequency from primary endosymbiosis of 11% in *Cyanophora* (Reyes-Prieto *et al.*, 2006) and 18% in *Arabidopsis thaliana* (Martin *et al.*, 2002).

E. gracilis shares a common ancestor with Kinetoplastida (*Trypanosoma*) (Adl *et al.*, 2005), which have never undergone secondary endosymbiosis (Rogers *et al.*, 2006). The *E. gracilis* plastid was acquired from a secondary endosymbiotic event although without traces of the secondary endosymbiont nucleus (McFadden, 2001). Judging from similarities in sequence analysis, *E. gracilis* appears to be a hybrid with a genome composed of *Euglena*-specific genes, Kinetoplastida-specific genes as well as common eukaryote

genes and genes acquired from the secondary endosymbiotic event most likely of photoautotrophic eukaryote origin having orthologues with only photoautotrophic eukaryotes (Henze *et al.*, 1995). About 14% of the genes have orthologues limited to photoautotrophic eukaryotes of which 25 % are related to photosynthesis while only 4% are derived from Kinetoplastida, most of which are unknown open reading frames (Ahmadinejad *et al.*, 2007).

Due to the horizontal gene transfer of the secondary endosymbiotic events it has been difficult to describe *Euglena* in a phylogenetic tree as the hybrid genome is derived from two ancestors (Dagan and Martin, 2006). Currently *Euglena* have been distributed in a monophyletic clade with Kinetoplastida although this only accounts for the Kinetoplastid ancestor and not the secondary endosymbiotic event (Figure 2) (Ahmadinejad *et al.*, 2007).

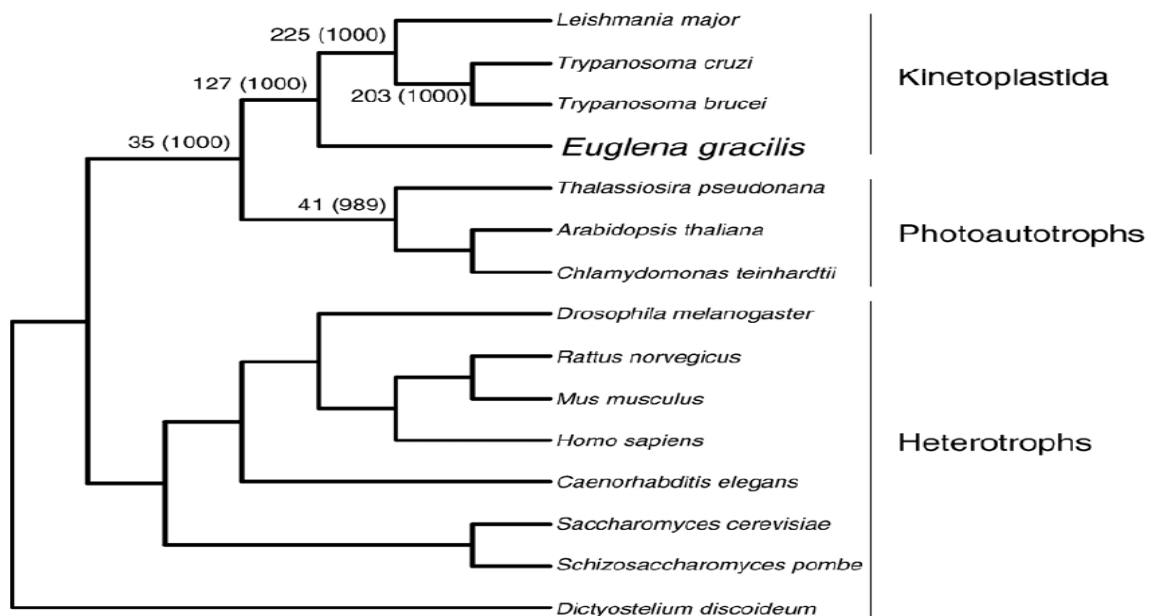


Figure 2: The phylogeny of *E. gracilis* where 259 globally distributed genes distributed in a phylogenetic tree grouped *E. gracilis* in a monophyletic clade with Kinetoplastids where the secondary endosymbiotic event is not represented (Ahmadinejad *et al.*, 2007).

1.3 Paramylon and paramylon synthesis

β -1,3-glucans are common in the natural world where they serve as structural frameworks, in the form of callose in plants (Samuels *et al.*, 1995), primary cell walls in fungi (Cabib *et al.*, 1987) and as exopolysaccharides in bacteria (Marchessault and Desandes., 1979). β -1,3-glucans exist as storage carbohydrates in the Phaeophyceae, Chrysophyceae and in Euglenophyceae (De Madariaga., 1992; Kiss

and Triemer., 1988; Kreeger and Van der Veer., 1970) as well as Heterokontophyta and Haptophyta (van den Hoek., 1995).

In most organisms that store β -1,3-glucans as carbohydrate reserve, the glucan polymers are non-granular as opposed to cytosolic granular *Euglenoid* paramylon (Barras and Stone., 1968). The highly crystalline substructure of granular paramylon is unique to *Euglena* (Marchessault and Deslandes, 1979), where the paramylon β -1,3-glucan chains are arranged in triple helical microfibrils arranged from a central region which connects the segments of the granule in a concentric pattern (Booy *et al.*, 1981; Deslandes and Marchessault, 1980; Marchessault and Deslandes., 1979).

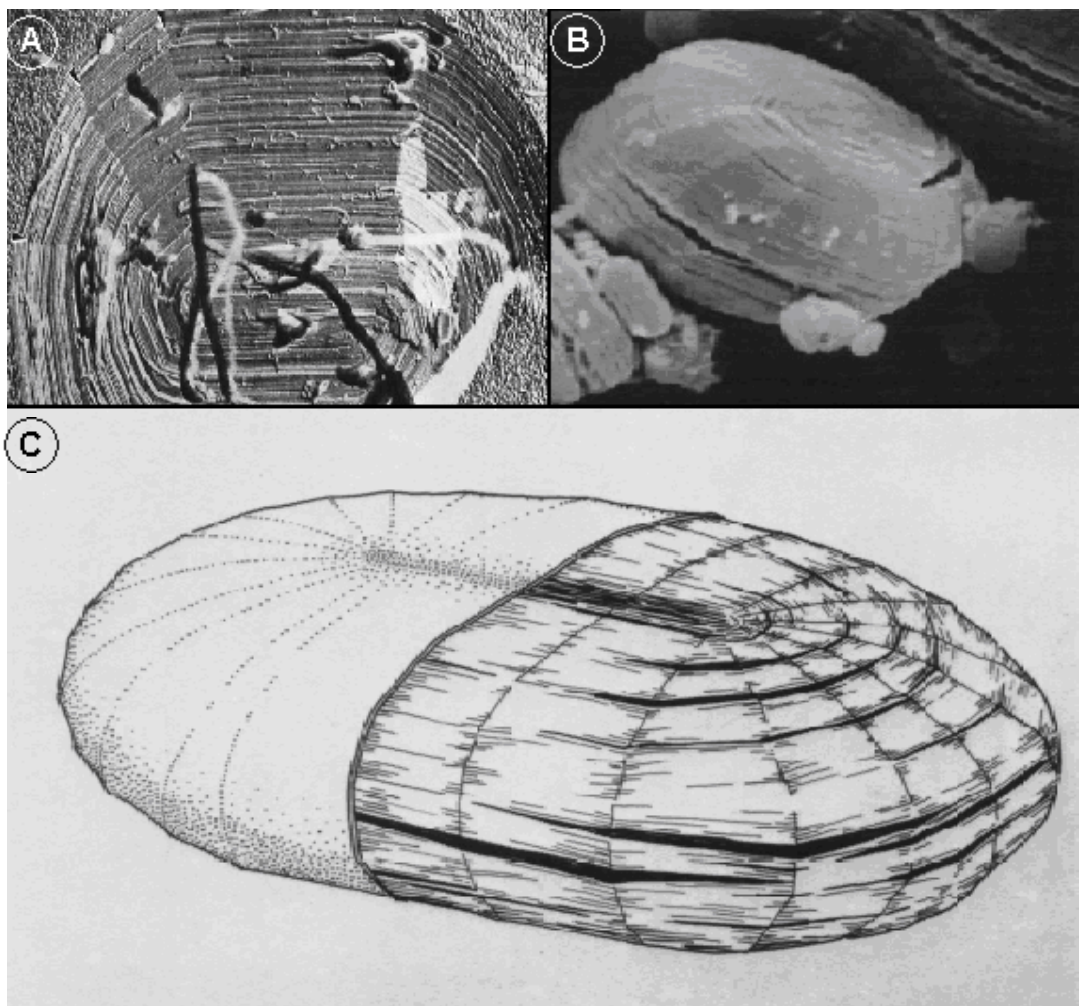


Figure 3: The structure of the paramylon granule. SEM of the paramylon granule from *E. gracilis* (A) (Kiss et al., 1987), SEM freeze fractured paramylon granule (B) (Kiss et al., 1987) and a diagrammatic representation of a paramylon granule surrounded by the membrane bilayer (C) (Kiss et al., 1987).

The triangular segments are arranged in three dimensional wedges, while the rectangular segments are block-shaped (Figure 3a). The microfibrils are arranged in complex segments with the layered mature microfibrils are arranged in the centre surrounded concentrically by maturing microfibrils (Figure 3a) (Barras and Stone., 1968; Kiss *et al.*, 1987; Marchessault and Deslandes, 1979) similar to the microfibril structure of cellulose (Gardner and Blackwell., 1971; Kato, 1981). The microfibrils are enveloped by a bilayer lipid membrane (Kiss *et al.*, 1988) which has large intramembraneous particles, which increase in density with granule synthesis (Figures 3b and 3c) (Kiss *et al.*, 1988). Paramylon was shown to be first synthesized as elementary microfibrils with low crystallinity (Kiss *et al.*, 1987, 1988a). As the granule matures the microfibrils assemble to form a highly ordered crystalline structure (Booy *et al.*, 1981; Kiss *et al.*, 1988a; Marchessault and Deslandes., 1979).

Paramylon granules are synthesized by membrane bound β -1,3-glycosyl transferases using UDP-glucose as the substrate (Marchessault and Deslandes., 1979) and are located cytosolically, surrounding pyrenoids in autotrophic cells and distributed freely at a 6-fold greater amount in heterotrophic cell cytosols (Briand and Calvayrac., 1980; Kiss *et al.*, 1986, 1988; Marechal and Goldberg., 1964; Tomos and Northcote, 1978). A protein primer is thought to be responsible for initial formation of the paramylon granule where the protein-glucan intermediate consists of glucan bound to an aglycone, partly by a phosphate ester bond, although the responsible protein has not yet been identified (Tomos and Northcote., 1978). A similar self glycosylating protein, a UDP-glucose glycosyl transferase, has been observed in plants (Quentmeier *et al.*, 1987).

The membrane surrounding the paramylon granule is distinctive as most other storage granules are not membrane bound (Kiss *et al.*, 1987). Freeze-etch fractionation of the membrane revealed two types of intramembraneous particles associated with the membrane, a large type of intramembraneous particles and smaller one. The membrane and intramembraneous particles are always associated with the granule and it is expected that new membranes are synthesized with the new granules as well as associated structural proteins (Kiss *et al.*, 1987). The frequency of the intramembraneous particles observed were related directly to paramylon synthesis and thought to be functional in paramylon synthesis (Kiss *et al.*, 1987). Other intramembraneous particles have been identified as transport proteins (Blaurock and Stoekenius, 1971; Henderson and Unwin, 1975) or related to cellulose synthesis (Mueller and Brown, 1980; Quader, 1986; Staehelin and Giddings, 1982).

The large and small intramembraneous particles both seem to be involved in paramylon synthesis where the large intramembraneous particles are more frequent during early stages of the granule synthesis and the smaller intramembraneous particle predominant in later stages of granule development. The larger intramembraneous particle may be the paramylon synthase complexes themselves while the smaller may be a transport or structural protein or involved in degradation (Kiss *et al.*, 1987). The high degree of

similarity between paramylon granules and the crystalline nature of cellulose suggest a similar mechanism of paramylon synthesis to cellulose synthesis where cellulose is synthesized by a membrane bound regulated multi-subunit enzyme complex (Delmer, 1999; Marchessault and Deslandes, 1979; Preston *et al.*, 1964).

1.4 Glycosyl transferases

Glycosyl transferases are enzymes that catalyze the covalent transfer of sugar residues between donor substrates and their acceptor molecules where the donor molecules are sugar phosphates, di-, oligo or polysaccharides and nucleotide diphosphate sugars and the acceptor molecule is usually a carbohydrate polysaccharide, lipid carrier or protein (Hundle *et al.*, 1992). Glycosyl transferases are sub-classified into two groups: processive glycosyl transferases such as cellulose, callose and chitin synthases that can transfer multiple sugar residues between donors and the acceptor polysaccharide while non-processive glycosyl transferases which transfer single sugar residues to their acceptor (Coutinho *et al.* 2003).

The vast majority of sequenced glycosyl transferases exist in the form of uncharacterized ORF's. Almost half of the sequenced glycosyl transferases are from sequenced genomes. Poly-specific families of glycosyl transferases have been included however glycosyl hydrolases and transglycosidases are not included in this system of classification as they are mechanistically and structurally unrelated (Henrissat, 1991; Henrissat and Bairoch, 1993; Henrissat and Bairoch, 1996). The greatest difficulties in glycosyl transferase characterization are due to multiple membrane-spanning domains and membrane associated activities (Laine, 1994). Substitutions of single amino acids may have significant impacts on glycosyl transferase activities such as substrate and product specificities where poly-specific families become difficult to predict with limited sequence information (Seto *et al.*, 1999). Structurally, glycosyl transferases can adopt one of two fold conformations but due to the multiple substrate donors and wide range of acceptor molecules, glycosyl transferases are complex and diverse in primary structure making functional prediction very difficult with limited sequence information (Coutinho *et al.* 2003).

Common domain organizations and interrelationships have been identified in several glycosyl transferases based on sequence comparison (Paulson *et al.*, 1989). Regions of distant sequence similarity have also been studied and identified (Atkinson and Long, 1992; Bulawa, 1992; De Angelis *et al.*, 1994; Debellé *et al.*, 1992; Dougherty and van de Rijn, 1994; Glucksmann *et al.*, 1993; Saxena *et al.*, 1994) suggesting that significant conserved domains are essential for the different glycosyl transferases. Very poor sequence similarities between glycosyl transferases have made global alignment impossible due to high divergence levels. However analysis aimed at secondary globular structures and transmembrane domains of the enzymes identified two conserved domains among glycosyl transferases where domain A, situated at the N terminal

half, is highly conserved and domain B, situated at the C terminal half, is conserved to a lesser extent. Glycosyl transferases consisting of both domains A and B appear to be processive glycosyl transferases (β -glucan synthases) opposed to non-processive glycosyl transferases having only domain A (Saxena *et al.*, 1995).

When glycosyl transferase amino acid sequences were searched for conserved amino acids having a side chain able to capably catalyze the acid-base transferase reaction, only two polar D residues were conserved among all the sequences (Zvelebil *et al.*, 1988). Domain A has been further analyzed (Gaboriaud, 1987; Lemesle-Varloot *et al.*, 1990) revealing vertically shaped clusters resembling β -strands alternating with α -helices where the D residues are invariably positioned in C terminal loops as expected with catalytic residues (Jacobson *et al.*, 1994; Zvelebil *et al.*, 1988). The hydrophobic clusters in domain B are more variable relative to domain A with only a single conserved D residue and the conserved QXXRW for cellulose synthases (Delmer, 1999; Saxena and Brown, 1999) or QXRRW for chitin synthases (Ruiz-Herrera *et al.*, 2002) sequence motifs presumed to be involved in the catalytic reaction (Hieta *et al.*, 1984). The bacterial β -1,3-glucan synthase genes share similar but not necessarily conserved domains where domain A shares the two D residues, one in the UDP-glucose binding domain as well as a KAG domain (YXXRXXXHAKAGNLTN) while domain B contains a TED domain with the third D residue and highly conserved QXXRW or QXRRW motifs (Stasionopoulos *et al.*, 1999).

Distant similarities between families point towards evolutionary divergence although all glycosyl transferase families are separated into two folding patterns, GT-A and GT-B (Bourne and Henrissat, 2001; Coutinho *et al.* 2003; Hu and Walker, 2002; Tarbouriech *et al.*, 2001; Ünlügil *et al.*, 2000) consisting of domain A, situated at the N terminal half and is highly conserved, and domain B, situated at the C terminal half, is conserved to a lesser extent (Saxena *et al.*, 1995). GT-A consists of two β - α - β domains which form continuous central sheets of at least eight β -strands. GT-B consists of two β - α - β domains associated weakly where ligand binding is associated with conformational changes with no significant sequence similarities between families (Wrabl and Grishin, 2001). Nucleotide binding occurs on the N-terminal of the GT-A enzymes and at the C-terminal of the GT-B enzymes while acceptor binding occurs on the opposite domain. The folds themselves do not determine the stereochemical reaction.

Glycosyl transferases are further separated into families GT1-65 where enzymes within the family although structurally related are functionally distinct. β -1,3-glucan synthases fall specifically under two gene families: GT2, a GT-A folding clan I inverting glycosyl transferase which has a wide range across all taxonomic kingdoms and consist of cellulose synthase, chitin synthase, β -N-acetylglucosaminyltransferase, β -N-acetylgalactosaminyltransferase, β -mannosyl transferase, dolichyl phosphate, hyaluronan synthase and β -1,3-glucan synthase. While GT48, an inverting glycosyl transferase not yet completely characterized,

belongs to plant and fungal kingdoms where only the β -1,3-glucan synthase makes up the family (Coutinho *et al.* 2003).

1.5 Partial purification of β -1,3-glucan synthases and identification of catalytic subunits

In plants, callose (β -1,3-glucan) and cellulose (β -1,4-glucan) are produced by an enzyme complex which is capable of synthesizing both polysaccharides under tight regulation (Delmer, 1999). The callose synthase genes, specifically the catalytic subunits, have been studied in various plants including carrot (*Daucus carota L*) (Lawson *et al.*, 1989), red beet (*Beta vulgaris*) (Frost *et al.*, 1990), barley (*Hordeum vulgare*) (Pedersen *et al.*, 1993), sunflower (*Lolium multiflorum*) (Bulone *et al.*, 1990), cotton (*Gossypium hirsutum*) (Delmer *et al.*, 1991; Li *et al.*, 1993; Li and Brown, 1993), soy bean (*Glycine max*) (Fink *et al.*, 1991) and pea (*Pisium sativa*) (Dhugga and Ray, 1994; 1998). Callose is deposited in localized periplasmic spaces after wounding, stress or pathogen attack where the accumulated β -1,3-glucan is broken down again when conditions return to normal and the specific stress is relieved through a rapid regulated turnover process (Currier, 1957; Eschrich, 1975; Fincher and Stone, 1981; Mueller and Maclachlan, 1983; Waterkeyn, 1981). Callose synthase and cellulose synthase are part of an ancient gene family, containing signature domains and conserved residues. A translated amino acid sequence could function as a synthase for both β -1,3- and β -1,4-glucan synthesis possibly modulated by changes in enzyme phosphorylation states, Ca^{2+} or possibly through a modifiable peptide subunit (Hill and Brew, 1975). β -1,3-glucan synthase activity in pea, soybean (Kauss and Jeblick, 1985; Krauss and Jeblick, 1986; Kauss *et al.* 1983) and cotton membranes (Delmer *et al.*, 1977; Delmer, 1983; Delmer *et al.*, 1985) is increased significantly in the presence of Ca^{2+} . The mechanism of Ca^{2+} activation is unknown, although it appears to be specific for β -1, 3-glucan synthases and not by Ca^{2+} activated kinases or calmodulin-mediated activation (Girard and Maclachlan, 1987). Callose is specifically activated by Ca^{2+} , however several chemical and physical treatments also induce callose formation. Physical damage to plant tissue may increase Ca^{2+} concentrations at the membranes or alternatively, leaked endogenous proteases may activate zymogen-like β -1, 3-glucan synthase activity. In plants it seems possible that either or both protease and Ca^{2+} may regulate callose formation (Girard and Maclachlan, 1987). This may also explain the results were active UDP-glucose binding protein subunits are often isolated together with slightly larger molecular weight inactive proteins (Girard and Maclachlan, 1987).

In fungi, β -1,3-glucans are integral parts of the cell wall (Baskin *et al.*, 1992) where some components exist as zymogens, where activity is regulated by proteolysis (Auer *et al.*, 1998; Carpita *et al.*, 1996). Genes encoding β -1,3-glucan synthases in fungi have been identified, sharing some identity with plant cellulose synthases. A gene encoding a large subunit of 200-230 kDa has also been identified which does not share

identity with plant or bacterial β -1,3-glucan synthases (Kelly *et al.*, 1996). A 170 kDa polypeptide was identified in mung bean (Shin and Brown, 1998), which showed some identity with the large fungal β -1,3-glucan synthase gene. Similarly a tobacco pollen tube β -1,3-glucan synthase showed sequence similarity and related molecular weight to the fungal gene product (Turner *et al.*, 1998). Amongst other examples, it seems likely that there is at least to some extent, some similarity between fungal and plant β -1,3-glucan synthases. The bacterial curdlan synthase gene encoding for a β -1,3-glucan synthase, was isolated and showed sequence similarities to cellulose synthase genes of bacteria and cellulose synthase genes of plants (Stasionopoulus *et al.*, 1999). This suggests that the cellulose synthase genes of plants may also be responsible for β -1,3-glucan synthesis with slight modifications in the catalytic site amino acid sequence (Campbell *et al.*, 1997; Legault *et al.*, 1995).

Fungal cell walls are complex structures consisting of chitins, mannans, β -1,3-glucans and β -1,6- glucans. Several other fungal β -1,3-glucan synthase related gene products have been identified although not catalytic subunits, are thought to be involved in cell wall biogenesis (Enderlin and Selitrennikoff, 1994; Hong *et al.*, 1994; Kasahara *et al.*, 1994a; Roemer and Bussey, 1991; Roemer *et al.*, 1993; Valdivieso *et al.*, 1991). Modification of the regulatory GTPase subunit of *Saccharomyces cerevisiae*, Rho1p, is essential for β -1,3-glucan synthesis. Prenylation of the protein subunit by the geranylgeranyl transferase is required for activation of the β -1,3-glucan synthase catalytic subunit (Arellano *et al.*, 1996; Diaz *et al.*, 1993; Inoue *et al.*, 1996; Inoue *et al.*, 1999; Kondoh *et al.*, 1997; Mazur and Baginsky, 1996; Quadota *et al.*, 1996) while the geranylgeranyl transferase in turn is regulated by cytosolic Ca^{2+} levels (Inoue *et al.*, 1999). KRE6 was reported to be the putative β -1,6-glucan synthase (Roemer and Bussey, 1991; Roemer *et al.*, 1993) which is closely inter-linked with the β -1,3-glucan synthase (Shematek and Cabib, 1980).

Several methods of extraction, purification and identification have been attempted to isolate β -1,3-glucan synthases (Table 1). The most popular methods, specifically for callose synthases, has been to partially purify the protein subunits usually through product entrapment and further identifying substrate binding subunits through photo affinity labeling assays or immunoprecipitation. Detergents such as CHAPS, W1 and digitonin have been used to solubilize the membrane bound proteins although loss of activity and instability after solubilization has made detection based on enzyme activity difficult. Attempts to purify the enzyme in fungi have also been relatively unsuccessful (Aswald *et al.*, 1993, 1994; Beauvais *et al.*, 1993) again having problems with enzyme stability during solubilization although *Saccharomyces cerevisiae* β -1,3-glucan synthase was partially purified through CHAPS solubilization and product entrapment (Inoue *et al.*,

Table 1: Methods of purification and identification of β -1,3-glucan synthase proteins in literature.

Source of β -1,3-glucan	Methods of purification and identification	Detergent	Catalytic subunits	Non-catalytic proteins	References
Plant					
<i>Beta vulgaris</i> (Red Beet)	photo affinity labeling UDP-glycosyl transferase activity	digitonin	57 kDa		Frost <i>et al.</i> , 1990
<i>Daucus carota</i> (Carrot)	photo affinity labeling UDP-glycosyl transferase activity	CHAPS	57 kDa	150 and 43 kDa	Lawson <i>et al.</i> , 1988
<i>Gossypium hirsutum</i> (Cotton)	photo affinity labeling	digitonin	52 kDa	44 kDa	Delmer <i>et al.</i> , 1991
	product entrapment photo affinity labeling	digitonin	37 kDa	32 and 66 kDa	Li <i>et al.</i> , 1993
<i>Hordeum vulgare</i> (Barley)	product entrapment sucrose density gradient centrifugation UDP-glycosyl transferase activity	digitonin	36 and 52 kDa	54, 60, 70 and 94 kDa	Pedersen <i>et al.</i> , 1993
<i>Pisium sativa</i> (Pea)	glycerol gradient centrifugation	digitonin	55 kDa		Dhugga and Ray, 1994; 1998
<i>Glycine max</i> (Soya bean)	polyethyleneglycol/dextran two phase separation sucrose density gradient centrifugation antibody affinity	digitonin	31 kDa		Fink, 1991
<i>Phaseolus vulgaris</i> (mung bean)	UDP-glycosyl transferase activity	digitonin	38 and 78 kDa	32, 54, 64 kDa	Kudlicka and Brown, 1997
<i>Lolium multiflorum</i> (Sunflower)	Immunoprecipitation photo affinity labeling	CHAPS	31 kDa	30, 54 and 58 kDa	Meikle <i>et al.</i> , 1991
Bacterial					
<i>Agrobacterium tumefaciens</i>	Functional complementation		59 kDa		Stasinopoulos <i>et al.</i> , 1999
Fungal					
<i>Saproglenia monoica</i>	Product entrapment UDP-glycosyl transferase activity assay	CHAPS	34, 48 and 50 kDa		Bulone <i>et al.</i> , 1990
<i>Saccharomyces cerevisiae</i>	Product entrapment	CHAPS	200 kDa		Inoue <i>et al.</i> , 1995
	Immunoprecipitation				
<i>Cordyceps militaris</i>	Cloned using sequence similarity		1981 aa		Ujita <i>et al.</i> , 2006
<i>Aspergillus fumigatus</i>	Cloned using sequence similarity	CHAPS	218 kDa	21.5, 100 and 160 kDa	Beauvais <i>et al.</i> , 2001
<i>Neurospora crassa</i>	Functional complementation			59 kDa	Enderlin and Selitrennikoff, 1994

1995). Functional complementation screening methods have been used more recently to identify β -1,3-glucan synthases (Enderlin and Selitrennikoff, 1994; Stasinopoulos *et al.*, 1999).

1.6 Paramylon synthase

Bäumer *et al.*, (2001) partially purified β -1,3-glycosyl transferase proteins from *E. gracilis* through separation by sucrose density centrifugation of proteins purified from the paramylon granular membrane through CHAPS solubilization. A 670 kDa enzyme complex was separated with UDP-glycosyltransferase activity, a molecular weight larger than that of most of the β -1,3-glycosyltransferase complexes in plant and bacterial cell membranes (Kudlicka and Brown., 1997; Li and Brown., 1993; Thelen and Delmer, 1986).

Bäumer *et al.* (2001) reported that paramylon synthesis is optimal with Ca^{2+} , spermidine (0.6 mM), CHAPS (0.15%) and a paramylon primer, similar to the requirements for cotton, carrot and green bean β -1,3-glucan synthases which require Ca^{2+} and a detergent, either digitonin or CHAPS (Hayashi *et al.*, 1987; Lawson *et al.*, 1989; McCormack *et al.*, 1997). Cytosolic Ca^{2+} concentrations over 200 nM signal tissue damage in higher plants triggering a wound repair response where callose β -1,3-glucan is synthesized while normal conditions (200 nM Ca^{2+}) induce cellulose β -1,4-glucan synthesis (Li *et al.*, 1993). In *E. gracilis* however no non specific glucan polymers were synthesized, only paramylon synthesis activity was affected by variations of Ca^{2+} (Bäumer *et al.*, 2001).

Bäumer *et al.* (2001) concluded that the paramylon synthase complex had a relatively high molecular weight (~670 kDa) with protein subunits similar and within range of those in previous studies (31 to 83 kDa) (Kudlicka and Brown., 1997) which could bind UDP-glucose as substrate under the positive influence of effectors.

1.7 Cytokine-related immunopotentiating activities of paramylon

β -1,3-glucans have drawn interest in the pharmacological field as biological moderators in human and animal systems specifically as immuno-potentiators, anti-inflammatory, antimicrobial, antitumor, hepatoprotective, antifibrotic, antidiabetic, hypoglycemic and anticholesterol activities (Brown and Gordon, 2003; Kojima *et al.*, 1986; Robbins and Seely, 1977; Sutherland, 1998; Trepel, 2004).

The macromolecular structures of the β -1,3-glucans are dependent on the sources of the glucans as well as the isolation procedure. *Euglena gracilis* paramylon is a water insoluble linear triple helical β -1,3-glucan (122 kDa), shown to have structural similarities to strong immunopotentiating β -1,3-glucans such as lentinan and schizophlan (Marchessault and Deslandes, 1979). Antitumor activity of native paramylon was reported to be relatively weaker than other pharmacologically active β -1,3-glucans, where only partial tumor regression was observed and most effectively at lower doses (1mg/kg/day x10) rather than higher doses (5mg/kg/day x10) (Quesada *et al.*, 1976). It has been suggested that paramylon could have greater potential due to its strong adjuvant activity other than antitumor responses, more specifically for the activities of paramylon on humoral antibody and cytokine production. The immunopotentiating effects of paramylon as an adjuvant has been observed at much higher doses (10 and 50 mg/kg) indicating that the adjuvant and antitumor activities follow different mechanisms (Kondo *et al.*, 1992). Paramylon administered to mice caused an increase in the levels of cytokine IL-6, which caused an differentiation of activated B cells to matured B cells (Hirano *et al.*, 1985; Muraguchi *et al.*, 1988), and high amounts of IL-1, which together with IL-6 caused an increase in T cell activation (Marchessault and Deslandes, 1979), as well as the induction of hematopoiesis for early precursor cells in bone marrow (Ikebuchi *et al.*, 1987; Wong *et al.*, 1988). The increase in IL-1, which appears higher than that of lentinan and schizophlan, have been shown to have multiple activities including T cell activation as well as IL-2 and IL-4 production and expression of IL-2 receptors (Ho *et al.*, 1987; Lowenthal *et al.*, 1980; Smith *et al.*, 1980). Paramylon appears to have greatest immunomodulating activities at higher doses and specifically on cytokine production. Anti-oxidative mechanisms were enhanced in the liver after paramylon administration, significantly reducing acute hepatic injury (Sugiyama *et al.*, 2009).

Several biologically active, specifically antimicrobial active derivatives of paramylon have been synthesized, positively charged derivatives including N,N-dimethylaminoethyl (DMAE), N,N-diethylaminoethyl (DEAE) and 2-hydroxy-3-trimethylammoniopropyl (HAP) were effective at increasing macrophage activities, specifically increasing NBT-reducing activities compared to negatively charged and neutral derivatives. However, carboxymethyl paramylon and paramylon sulfate did not significantly enhance specific macrophage antimicrobial activities (Sakagami *et al.*, 1989; 1991), only paramylon sulfate has effective anti-HIV activities against HIV-1nTLVmm, HIV-1AO12B and HIV-2ROD in PBMC and MT-4 cells (Baba *et al.* 1988; De Clercq, 1986; Ito *et al.* 1987; Mizumoto *et al.*, 1988; Nakashima *et al.*, 1987; Ueno and Kuno, 1987). The anti-HIV activities of paramylon sulfate was completely dependent on molecular weight and degree of sulfation (Baba *et al.*, 1990; Daniel *et al.*, 1985; Itoh *et al.*, 1990; Witvrouw *et al.*, 1991), where paramylon sulfate completely inhibited HIV binding to cells in a dosage range compatible with *in vivo* administration

(Koizumi *et al.*, 1993). Although paramylon sulfate had a lower inhibition of HIV binding compared to dextran sulfate (Baba *et al.*, 1988; Mitsuya *et al.*, 1988; Nakashima *et al.*, 1989; Schols *et al.*, 1989), it was still higher than the inhibition of HIV binding by polyphenolic compounds (Manabe *et al.*, 1992; Nakashima *et al.*, 1992a; 1992b). The synergistic anti-HIV effect was increased when combinations of azidothymidine (AZT) and paramylon sulfate were optimized, reducing therapeutic dose requirements and anti-HIV activity (Koizumi *et al.*, 1993). Antiviral activities are only specific for HIV (Fukuchi *et al.*, 1989; Harada *et al.*, 1991)

1.8 Aims and experimental approach

The aim of this study was to identify genes involved in the synthesis of paramylon, specifically the catalytic β -1,3-glycosyl transferase genes. In other studies, several methods have been used in attempt to isolate and characterize the paramylon synthase genes including functional screening of cDNA libraries, where various detection methods were used (van der Merwe, 2007) as well as protein purification approaches (Bäumer *et al.*, 2001; van der Merwe, 2007). In this study, several strategies have been followed, one of the strategies has been to construct and screen a cDNA library using a wider range of screens than in previous studies.

A *Euglena gracilis* cDNA lambda phage library was constructed which expressed gene products by lysing XL1- blue *E. coli* or expressing the cDNA library inserted into pBluescript plasmids in *E. coli* colonies. Aniline blue served as a functional screen as it specifically stains β -1,3-glucans although this strategy relies on the proteins being actively expressed in *E. coli*, as well as the stain being active within the bacterial colonies. Another technique was to use the ZAP cDNA synthesis kit in combination with the Pico Blue immunoscreening protocol where the phage library was expressed in the form of Lambda phage containing pBluescript vectors containing *E. gracilis* cDNA inserts. Lysis of bacterial lawns by the phage library would express the cDNA gene products. The gene products would be expressed and blotted onto a nitrocellulose membrane and blocked. These membranes would be screened either using radio-labeled substrate in the form of [14 C] UDP-glucose where the substrate binding protein subunits (Bäumer *et al.*, 2001; van der Merwe, 2007) would be screened. Another approach was to screen phage libraries with an antibody raised against the proteins associated to the paramylon granule as this would not be dependent on functional activity for detection. The cDNA library was amplified and purified by mass excision in the form of pBluescript. This allowed for expression of the gene products directly in *E. coli* systems where they were screened by lysing the bacterial colonies and detecting the gene products using the antibody. The mass

excision was also screened using primers to amplify DNA based on consensus sequences derived from β -1,3-glucan synthases from several organisms as well as degenerate and non degenerate primers designed from a peptide sequence isolated from purified paramylon synthase proteins.

Functional and structural instability of the isolated proteins have caused difficulty in initial attempts to purify these proteins in other systems (Delmer, 1999). However, a relative degree of success has lead to isolation of catalytic subunits through protein purification in conjunction with the available sequence information, where peptide sequences isolated were compared to pre-existing sequenced genomes. Due to the relative lack of sequence information available for *E. gracilis*, protein purification methods to deduce gene information rely greatly on a degree of identity to already characterized reference genes.

The protein purification strategy adopted for this study aimed at purifying proteins associated with the paramylon granule membrane. Previous studies were partially successful at partially purifying two peptide subunits with substrate affinity comparable to those found in plants (Bäumer *et al.*, 2001; van der Merwe, 2007). This study has followed on from previous methods in order to partially purify and derive sequence information from peptides associated with the granule membrane.

A transcriptomic approach was also followed to isolate transcript sequence information from *E. gracilis*. As the genome is made up of a chimeric collection of genes inherited from a combination of a heterotrophic kinetoplastid and the chloroplast of a photosynthetic eukaryote obtained during a secondary endosymbiotic event, making *E. gracilis* taxonomically complex with respect to both morphology and genomics as no comparable reference genome exists. The chloroplast of *E. gracilis* has already been sequenced along with several ribosomal and EST sequencing projects already (Ahmadinejad *et al.* 2007) in order to gather sequence information specifically for taxonomic and expression profile studies. During this study, an un-annotated genome was sequenced and released by another group in a taxonomic study, which was then also used in this study.

In this study, the transcriptome was sequenced under conditions optimal for paramylon synthesis using the genome sequencer FLX-system specifically to identify genes involved in paramylon synthesis. Previous attempts to sequence *E. gracilis* ESTs have not identified any β -1,3-glycosyl transferases although short glycosyl transferases GT-2 sequences were identified on Genbank as well as β -1,3-glycosyl hydrolases.

Materials and Methods

2. Materials

2.1 Chemicals and kits

Chemicals and kits used were obtained from Sigma (St. Louis Missouri, USA), Roche Diagnostics (Mannheim, Germany), Promega (Madison, Wisconsin, USA), Stratagene (La Jolla, California, USA), Invitrogen (Carlsbad, California, USA), CalBiochem (Merck Biosciences, Darmstadt, Germany), BioRad Laboratories GmbH (Munich, Germany), Fermentas, Amersham biosciences (Uppsala, Sweden) GE healthcare (Uppsala, Sweden).

Paramylon synthase specific primary antibodies were produced by the Department of Biochemistry, University of Stellenbosch, South Africa. Secondary antibodies were obtained from Sigma (St. Louis Missouri, USA).

Euglena gracilis Z Krebs strain SAG 1224-5/25 obtained from Algensammlung, Göttingen, Germany.

SOLR™ strain (Stratagene) $e14^-$ ($McrA^-$) $\Delta(mcrCB-hsdSMR-mrr)$ 171 sbC $recB$ $recJ$ $uvrC$ $umuC::Tn5$ (Kan^r) $lacgyr$ A96 $relA1$ $thi-1$ $endA1$ λ^R [F' $proAB$ $lac^qZ\Delta M15$] Su^- (nonsuppressing)

XL1- Blue MRF' strain (Stratagene) $\Delta(McrA)$ 183 $\Delta(mcrCB-hsdSMR-mrr)$ 173 $endA1$ $supE44$ $thi-1$ $recA1$ $gyrA96$ $relA1$ lac [F' $proAB$ $lac^qZ\Delta M15$ $Tn10$ (TET^r)]

DH5 α strain (Promega): $F'/endA1$ $hsdR17(r_k^- m_k^+)$ $supE44$ $thi-1$ $recA1$ $gyrA$ (Na^r) $relA1$ $\Delta(lacZYA-argF)$ $deoR$ ($\Phi80dlac\Delta(lacZ)M15$)

BL21 strain (Invitrogen): *E. coli* B F^- , $ompT$, $hsdS$ (r_B^- , m_B^-), gal , dcm .

2.2 Primers

All primers were obtained from IDT (USA) and were designed using NCBI primer design. Primers were designed based on β -1,3-glucan synthase sequences (Table 2) which were aligned using EBI/t-coffee where multiple alignments and conserved domains were identified. Forward and reverse degenerate primers were designed based on conserved amino acid sequences (Table 3) according to the codon

usage efficiency of *E. gracilis* (Nakamura *et al.*, 2000). The number assigned to the primers are based on the location of the sequence relative to the *Candida albicans* β -1,3-glucan synthase.

Table 2: Conserved β -1,3-glucan synthase sequences used to design primers.

Source of β -1,3-glucan synthase	NCBI Accession number
<i>Candida albicans</i>	ref XP_721429.1
<i>Lolium multiflorum</i>	gb AAQ17229.1
<i>Schizosaccharomyces pombe</i>	ref NP_594766.1
<i>Saccharomyces cerevisiae</i>	gb AAR86936.1
<i>Scedosporium prolificans</i>	gb ABY53595.1
<i>Arabidopsis thaliana</i>	ref NP_567278.1

Table 3: Degenerate primers designed from β -1,3-glucan synthase conserved domains.

Primer CD1	1075-1125
Fw:	5'- AAC CAG GAY AAC YAC HTI GAG GAG -3'
Rev:	5'- CTC CTC NAI GTR GTT RTC CTG GTT -3'
Primer CD2	943-963
Fw:	5'- TGG GCN AGC TWS CGS KSY CAG CAN YTS -3'
Rev:	5'- SAR NGT CTG RSM SCG SWA GCT NGC CCA -3'
Primer CD3	1191-1204
Fw:	5'- HTN CAC TAC GGN CAC CCN GAY -3'
Rev:	5'- RTC NGG GTG NCC GTA GTG NAI -3'
Primer CD4	1210-1232
Fw:	5'- GAG GAY ATY TYC GCA GGA YTS -3'
Rev:	5'- SAR TCC TGC GRA RAT RTC CTC -3'
Primer CD5	1421-1437
Fw	5'- GAG TGC YTY TGC TWC ATW TWC MAC -3'
Rev	5'- GTK GWA WAT GWA GCA RAR GCA CTC -3'
Primer CD6	321-332
Fw	5'- SCC GAG TGC CTG TGC TWC ATY -3'
Primer CD7	392-400
Fw	5'- TAC GAY GAY IKY AAC SAG HWS TTC TGG -3'
Rev	5'- CCA GAA SWD CTS GTT RMI TTC RTC GTA -3'

2.3 Primers designed from an isolated *E. gracilis* β -glucan synthase sequence

A short peptide sequence (KMCFTQGYLEFSARL) was predicted from a previous study from a ~50 kDa protein subunit purified and solubilized from the paramylon granular membrane through SDS-PAGE purification, tryptic digestion and LCMSMS (van der Merwe, 2007). The peptide sequence showed complete homology to a *KRE6* β -1,6-glucan synthases from *Debaromyces hanseii* and partial homology identity to *KRE6* β -1,6-glucan synthases from *Saccharomyces cerevisiae* and *Candida albicans*. Forward and reverse primers were designed (Table 4) based on the predicted sequence using codon usage efficiency of *E. gracilis* (Nakamura *et al.*, 2000).

The *Debaromyces hanseii SKN1* (XP 002770128.1), *Candida tropicalis KRE6* (XP 002547982.1) and *SKN1* (XP 002547983.1) β -1,6-glucan synthases have conserved WN amino acids on the N-terminal of the KMCFTQGYLEFSARL sequence of the glucan synthase peptide which were included in the primer design as they reduced degeneracy of the primers.

Table 4: Primers based on the *D. hanseii* glucan synthase peptide.

Primer	β -glucan synthase primer (MCFTQGY)
Fw non deg	5'- ATG TGC TTT ACC CAA GGG TAC T -3'
FW deg	5'- CG ATG TGC TTY CAN CAR GG -3'
Primer	β -glucan synthase primer (WNKMCFT)
Fw non deg	5'- TGG CAG AAG ATG TGC TTC AC -3'
FW deg	5'- TGG CAR AAR ATG TGY TTY AC -3'

2.4 Web – based programs

Table 5: Website-based bioinformatics tools.

Database/ Program	http site	Reference
Bioedit	http://www.mbio.ncsu.edu/BioEdit/	Hall, 1999
NCBI BLAST	www.ncbi.nlm.nih.gov/Blast/	Altschul <i>et al.</i> , 1990
CLUSTALW	www2.ebi.ac.uk/Clustalw	Thompson <i>et al.</i> , 1994
<i>Euglena</i> EST Database	tbestdb.bcm.umontreal.ca/	O'Brien <i>et al.</i> , 2007
EBI/ t-coffee	http://www.ebi.ac.uk/Tools/t-coffee/	Notredame <i>et al.</i> , 2000
Codon usage table	http://www.kazusa.or.jp/codon	Nakamura <i>et al.</i> , 1997
Mascot	www.matrixscience.com/	Hirosawa <i>et al.</i> , 1993
GeneMark	http://exon.biology.gatech.edu/	Besemer and Borodovsky, 1999
CLC genomics workbench	http://www.clcbio.com/	Droege and Hill, 2008

3. Methods

3.1.1 Construction and screening of the cDNA library

3.1.1.1 Growth conditions of *Euglena gracilis*

Euglena gracilis was grown under heterotrophic conditions, at 22°C, on media consisting of 10g tryptone, 4g yeast extract and 30g D-glucose per L. Liquid cultures were grown on a rotational shaker at 130 rpm and additional glucose was supplemented to 2% (m/v) once a week. Solid media contained 1.5% agar and were transferred to new plates every month.

3.1.1.2 Total RNA extraction

RNA was extracted from liquid cultures. Cells were centrifuged at 4,200g for 10min. Cells were homogenized in liquid nitrogen and added to 5 volumes extraction buffer preheated at 65°C consisting of 2% CTAB, 2% PVP, 100mM Tris-HCl pH 8.0, 25mM EDTA, 2M NaCl, 0.05% spermadine trihydrochloride and 3% β-mercaptoethanol. Extracts were vortexed immediately and incubated at 65°C for 5 min. After incubation the extracts were centrifuged at 8000g and the supernatant was transferred to a new tube. One volume of chloroform: isoamyl alcohol (24:1) was added to the supernatant, vortexed for 30 seconds and centrifuged again at 6000g for 15min. The aqueous layer was pipetted into a new tube and this step was repeated. LiCl was added to a final concentration of 2M and the RNA was precipitated overnight at 4°C. The precipitated RNA pellet was collected by centrifugation at 10000g for 60min and washed in 500ul 70% ethanol. The washed RNA was centrifuged at 10000g for 15min and dried. The RNA was resuspended in distilled water.

3.1.1.3 Isolation and quantification of mRNA

Messenger RNA was purified from total RNA using the Genelute™ mRNA miniprep kit (Sigma MRN10KT 0461K6879) according to product guidelines. All RNA was quantified using a nanodrop 2000 (Thermo Fischer Scientific Inc., USA). Concentrations were measured at 260nm and RNA quality was calculated as the ratio between 260 and 280nm and samples were tested using 1% agarose gel electrophoresis.

3.1.1.4 cDNA synthesis

First strand cDNA synthesis from isolated mRNA was performed using the Revertaid™ H Minus First strand cDNA synthesis Kit (Fermentas) according to product guidelines. Second strand synthesis of cDNA was performed using the second strand cDNA synthesis kit (Stratagene, #200401-5) according to the product guidelines. Purified second strand cDNA was quantified using a nanodrop 2000 (Thermo Fischer Scientific Inc., USA). The quality of the cDNA was determined by agarose gel electrophoresis. The purified second strand cDNA was quantified using a nanodrop 2000 (Thermo Fischer Scientific Inc., USA). The quality of the cDNA was determined by agarose gel electrophoresis.

3.1.1.5 cDNA library construction and in vitro packaging of the phage library

The cDNA library was constructed using the Zap cDNA Synthesis kit (Stratagene), *E. gracilis* total RNA was extracted and Poly(A)⁺ RNA was purified using the Genelute™ mRNA miniprep kit (Sigma). First strand cDNA synthesis, second strand cDNA synthesis, size fractionation, ligation of the cDNA insert into the Uni-Zap XR vector and all related steps were performed according to the product guidelines. The cDNA Uni- Zap XR ligated vector was packaged using the Gigapack III Gold cloning kit (Stratagene) according to the product guidelines.

3.1.1.6 Amplification of the cDNA primary library

XL1-Blue MRF' cells (Stratagene) were used for titrating and the construction of the amplified library following the instructions of the Gigapack III Gold cloning kit (Stratagene) manual. The titer was calculated (186 plaques per plate 1:10 dilution). The amplified library was collected and stored in SM buffer (100mM NaCl, 50mM Tris-HCL pH. 7.5, 10mM MgSO₄) in 2ml aliquots at -80°C.

3.1.1.7 Mass excision of the amplified library

A mass excision of the amplified library was performed using XL1-Blue MRF' (Stratagene) and SOLR™ (Stratagene) bacterial strains and the ExAssist helper phage according to the Gigapack III Gold cloning kit (Stratagene) manual. The mass excision was performed according to the titer of the amplified library. Following the mass excision the phagemid library was purified using the Nucleobond® AX nucleic acid purification kit (Macherey-Nagel, Germany).

3.1.1.8 Single clone excision

Plaques of interest were cored out from agar plates. And transferred to new tubes and stored in SM buffer (100mM NaCl, 50MM Tris-HCl pH 7.5 and 10mM MgSO₄) and 20μl chloroform. Phages were extracted and the pBluescript phagemid vectors containing ligated inserts were purified according to the Gigapack III Gold cloning kit (Stratagene) guidelines. Phagemid vectors were transformed into SOLR™ (Stratagene) bacterial cells. Plasmid DNA was prepared from colonies and the plasmids were sequenced.

3.1.2 Screening of the phage library

3.1.2.1 Infection and Expression of the phage library

The phage library was screened using the PicoBlue™ Immunoscreening Kit (Stratagene) according to the manufacturer's guidelines. Cell cultures of XL1-Blue MRF' (Stratagene) were grown overnight in NZY broth (5g NaCl, 2g MgSO₄·7H₂O, 5g yeast extract, 10g NZ amine supplemented with 2% maltose) and resuspended in 10mM MgSO₄ and 2% maltose to an OD₆₀₀ of 0.5. Known titers of Lambda Zap phage from the amplified library were used to infect the 200μl aliquots of cell cultures. The infected cell cultures were incubated at 37°C for 15min and added to NZY top medium (0.7% agarose) pre warmed to 48°C and poured onto pre warmed NZY plates. The plates were incubated upside down for three and a half hours at 37°C until small phages appeared. Nitrocellulose membranes (Millipore), pre soaked and 10mM IPTG and dried, were placed on the lysed bacterial lawns and the positions of the membranes were marked using a sterile needle. The plates were incubated with the membranes for another four hours at 37° and then briefly at 4°C. The membranes were lifted for further screening and the plates were stored at 4°C.

3.1.2.2 Screening of the phage library using an antibody

The membranes were screened using a polyclonal primary antibody raised in rabbits (University of Stellenbosch, South Africa). The antibody was raised against proteins associated (bound) to the paramylon granule.

Membranes lifted from phage infected plates were washed in TBST (20mM Tris-HCl pH 7.5, 150mM NaCl and 0.05% Tween 20) and blocked in blocking solution consisting of TBST and 1% BSA (Bovine Albumen (Fraction V), Roche) for two hours. Membranes were washed and blocked in square containers with several membranes per container and 8ml/membrane blocking solution. The primary antibody was added to fresh blocking buffer in a 1:500 dilution and incubated overnight. The membranes were washed in TBST 5 times and incubated for one hour in TBST with the secondary goat anti rabbit – alkaline phosphatase antibody in a 1:20000 dilution. The membranes were washed five times in TBST and once in TBS (TBST without Tween 20) and incubated in a solution of NBT/BCIP Ready-to-use tablets (Roche) until colour development was observed. The membranes were washed extensively with water to stop the colour development reaction.

The excised phages were vortexed and centrifuged at 10000g for 5 min. The phages of interest were screened again by infecting fresh XL1-Blue MRF' (Stratagene) cells, repeating the screening process for secondary and tertiary screens using 2 µl of the isolated plaque for infection. Phages that appeared to produce signal during secondary and tertiary screens were extracted by single clone excision.

3.1.2.3 Screening of the phage library using radio-labeled substrate

Nitrocellulose membranes (Sigma-Aldrich) were screened using [¹⁴C] UDP-glucose. Membranes lifted from phage infected plates were washed in TBST and blocked in buffer in square containers containing approximately 8ml buffer per membrane. Membranes were blocked in 4% BSA-TBST solution for two hours. The membranes were incubated in 250mM Tris-HCl pH 7.5 and 10µl [¹⁴C] UDP-glucose for 1 hour at room temperature with gentle agitation. The membranes were dried and exposed to a Super resolution Cyclone Phosphor screen (Packard). After seven days the hybridization was detected with a Cyclone TM Storage Phosphor system phosphor imager (Packard instrument Co., Meriden, USA). Radioactive signals which corresponded to phages were determined and the phages were excised and stored in 500µl SM buffer (100mM NaCl, 50MM Tris-HCl pH 7.5 and 10mM MgSO₄) and 20µl chloroform.

The excised phages were vortexed and centrifuged at 10000g for 5 min. The phages of interest were screened again by infecting fresh XL1-Blue MRF' (Stratagene) cells, repeating the screening process for secondary and tertiary screens using 2 µl of the isolated plaque for infection. Phages that appeared to produce signal during secondary and tertiary screens were extracted by single clone excision.

3.1.2.4 Functional screening of the mass excision library using aniline blue

The mass excised pBluescript vector based phage library was transformed into BL21 *E. coli* competent cells. Approximately 100ng DNA of the plasmid library was used to transform 1 ml of BL21 competent cells per large plate made up of LB solid medium with ampicillin and 2 mM IPTG added. The transformed cells were added to 10ml/ plate LB medium containing 0.7% agarose and ampicillin and poured onto the plates. The embedded transformed colonies were misted using a spray bottle with aniline blue solution (0.0025% aniline blue (Sigma) dissolved in 0.1M K₃PO₄ pH 12) and incubated open at room temperature for 30 min on a laminar flow bench. The plates were washed with 1 M NaCl to destain the colonies. The plates were viewed under ultraviolet light to identify possible positive colonies staining yellow.

3.1.2.5 Screening of the mass excision library using cell lysis

The immunological screening of the phage expression library was performed with transformed bacterial colonies according to the Sambrook and Russel, 2001. Nitrocellulose membranes (SIGMA, Millipore #N-7395) were placed on LB solid medium containing ampicillin. A culture of SOLAR cells (Stratagene) were inoculated from a single colony and grown in LB liquid media containing ampicillin until the OD₆₀₀ reached 0.5. A tittered amount of phage was used to inoculate the cell culture and incubated at 37°C for 15 min. The phage inoculated cell cultures were aliquoted onto the nitrocellulose membranes and spread evenly using a sterile glass rod. The plates were incubated for approximately 12 hours at 37°C until small bacterial colonies appeared. Replicate nitrocellulose membranes were placed on top of the primary membranes and the orientations were marked using a sterile needle. The replicate membranes were placed onto new LB plates containing ampicillin and 2 mM IPTG (isopropylthio-β-D- thiogalactopyranoside) and incubated for approximately 8 hours until the replicate bacterial colonies formed. The colonies on the nitrocellulose membranes were then fixed by exposing the membranes to chloroform fumes for 20min in a closed container. The bacterial colonies on the membranes were then lysed by incubating the membranes for 16 hours at room temperature on a shaker in blocking solution (TBST containing 4% BSA) with 1mg/ml 0.1% lysozyme and 0.05% [w/v] sodium azide added. Membranes were washed and blocked in large round Petri dishes with two membranes per container and 10 ml/membrane solution. The primary antibody was added to fresh blocking buffer in a 1:500 dilution and incubated overnight. The membranes were washed in TBST 5 times and incubated for one hour in TBST with the secondary goat anti rabbit – alkaline phosphatase antibody in a 1:20000 dilution. The membranes were washed five times in TBST

and once in TBS (TBST without Tween 20) and incubated in a solution of NBT/BCIP Ready-to-use tablets (Roche) until colour development was observed. The membranes were washed extensively with water to stop the colour development reaction.

3.1.3 cDNA synthesis for PCR screening and molecular cloning

3.1.3.1 Design of PCR primers

Primers were designed (Section 2.4) from short peptides sequences isolated from paramylon synthase peptides purified from *E. gracilis* (van der Merwe, 2007). Degenerate and non degenerate primers were designed and used to screen both the mass excision library as well as cDNA for sequence. Primers were also designed based on conserved domains of β -1,3-glycosyl transferase shared by a number of organisms.

Degenerate and non-degenerate primers were designed based on a peptide sequence isolated from a peptide partially purified from *E. gracilis* showing UDP-glucose binding affinity which separated at approximately 37 kDa on SDS PAGE. The peptide was digested with trypsin and the sequence was obtained by LCMS (van der Merwe, 2007). These primers were used as well as new primers designed from the same sequence for this study.

3.1.3.2 PCR conditions

PCR conditions varied for T_m and elongation times depending on the primer pairs used. PCR conditions for cDNA templates were denaturation at 94°C for 2 min, 35 amplification cycles of 94°C for 45s, 30s at 55 or 60°C and 60s or 90s at 72°C and final elongation for 7min at 72°C. PCR conditions for mass excision plasmid templates varied for T_m and elongation times. As pBluescript formed the vector backbone of the mass excision library, the T3 primer (5' – ATTAACCCTCACTAAAGGGA – 3') was used as an anchored primer, as the direction of cDNA insertion was in this direction (Zap cDNA Synthesis kit guidelines, Stratagene), where the conserved domain reverse primers and designed peptide primers were paired with the T3 primer, where the T7 primer (5' – TAATACGACTCACTATAGGG – 3') was paired with the conserved domain forward primers. PCR conditions were 2 min denaturation at 94°C, 35 amplification cycles of 94°C for 30s, 55°C for 30s and 72°C for 90s and final elongation at 72°C for 7min.

3.1.3.3 DNA purification, cloning, transformation and sequencing

PCR products were separated using agarose gel electrophoresis, stained with ethidium bromide and visualized using the Alpha imager. All PCR products were excised and purified using the QIAquick® Gel Extraction kit (QIAGEN) according to the product guide lines. Purified PCR products were ligated using the Clonejet™ PCR Cloning Kit (Fermentas) according to the product guidelines. All bacterial transformations were performed according to using the heat shock method (Maniatis protocol). Colonies that were selected were screened using the CloneJET™ kit's primer pair (forward sequencing primer: 5'-CGACTCACTATAGGGAGAGCGGC-3' and reverse sequencing primer 5'-AAGAACATCGATTTTCCATGGCAG-3'. Vectors containing inserts of the correct sizes were excised and sequenced using the T7 primer as described. All sequencing analysis was done at the Central Analytical Facility, University of Stellenbosch, using the ABI Prism 3730 Genetic Analyser using bigDye terminator cycle sequencing ready reaction.

3.2 Partial purification of paramylon synthase proteins

3.2.1.1 Protein extraction

Crude protein was extracted according to a modified method of Bäumer *et al.* (2001). Heterotrophic cell cultures were centrifuged at 4,200g for 10min in a Sorval centrifuge (Sorval RC5C Plus, Kendra Lab products, USA). The pellets were washed twice with ice cold distilled water and resuspended in buffer A (50mM Tris-HCl pH 7.4, 250mM sucrose, 3mM EDTA, 1mM dithiotreitol (DTT, Roche) and 1 tablet Complete™ protease inhibitor (Roche) per 50ml extraction buffer). Cells were disrupted by sonication using a virsonic 60 sonicator (The Viritis Co. Inc., Gardiner, New York, USA) at maximum for four 15 sec bursts and 30 seconds intervals on ice. The paramylon granules was centrifuged at 1500g for 5min at 4°C, washed four times in buffer A and resuspended in buffer B (25mM Tris-HCl pH 7.4, 20% (w/v) sucrose and 1mM DTT). The protein was solubilized from the paramylon granules by adding a solution of the detergent CHAPS (3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulfonate, Sigma) drop wise to a final concentration of 0.15% (w/v). The proteins associated to the paramylon granule were solubilized overnight with gentle stirring on ice. The paramylon granules were separated from the disassociated protein by centrifugation at 2000g for 5min at 4°C. The partially purified protein extracts were stored at -20°C.

3.2.1.2 Dialysis of the crude protein extract

Partially purified protein samples were desalted using 10,000 MW Slide-A-Lyzer® Dialysis Products (Pierce, #66810). Protein samples were loaded into the slide according to manufacturer's guidelines. The slide was desalted in 2l of an equal strength buffer (25 mM Tris-HCl pH (7.4) and 1mM DTT) overnight with gentle stirring until the protein sample became isotonic with the buffer.

3.2.1.3 Protein quantification

Protein was quantified according to the method of Bradford (1976) using the BioRad protein assay reagent (BioRad, #500-0006). The protein concentration was calibrated using a Bovine Serum Albumin (BSA) standard curve. Protein was quantified by measuring the absorbance at 595 nm using a μ Quant microplate scanning spectrophotometer (Bio-Tek Instruments, Winooski, Vermont, USA). Protein samples partially purified with CHAPS were quantified using the BCA™ Protein Assay Kit-Reducing Agent Compatible (Pierce, Cat. 23250, Rockford Illinois, USA) according to the product guidelines. Absorbance was measured at 562 nm using a μ Quant microplate scanning spectrophotometer (Bio-Tek Instruments, Winooski, Vermont, USA) calibrated against a BSA standard curve.

3.2.2 FPLC, PAGE separation, product entrapment and immunoblotting of partially purified paramylon synthase proteins

3.2.2.1 FPLC of paramylon synthase proteins

The anti-paramylon synthase antibody was purified using FPLC and coupled to activated CNBr-sepharose. Dialysed protein extract was purified by affinity binding FPLC of the antibody coupled to activate CNBr-sepharose.

Polyclonal antibody serum (500 μ l) was desalted using an Amicon Ultra 10 000 kDa MWCO centrifuge column and purified through the HiTRAP Protein A with sample size at 1ml, where 20mM sodium phosphate, pH 7.0 was used as the binding buffer where 0.1 M citric acid pH 4.5 was used for elution. The eluted purified polyclonal antibody was resuspended in CNBr-activated Sepharose 4B coupling buffer (0.1 M NaHCO₃, pH 8.3, 0.5 M NaCl) and coupled to CNBr-activated Sepharose 4B (GE healthcare) for 2 hours at room temperature and washed with three cycles of 0.1 M acetic acid/

sodium acetate, pH 4.0 containing 0.5 M NaCl alternating with 0.1 M Tris-HCl, pH 8 containing 0.5 M NaCl. Dialyzed CHAPS solubilized protein samples (Section 3.2.1.2.) were equilibrated in the binding buffer and desalted using a PD-10 column (GE Healthcare).

Affinity binding of dialysed protein extract was performed using an ACTA prime FPLC with the HiTRAP Protein A sample size at 1ml, where 20mM sodium phosphate, pH 7.0 was used as the binding buffer where 0.1 M citric acid pH 4.5 was used for elution. Dialysed protein (2.5 ml, ~5000 mg) was injected into the antibody coupled sepharose column where proteins bound to the CNBr coupled antibodies in binding buffer (75mM Tris-HCl, pH 7.4) and then eluted with elution buffer (0.1 M glycine-HCl pH 2.7, 0.5 M NaCl).

3.2.2.2 Detection of in gel paramylon synthase activity

Partially purified native proteins disassociated from the paramylon granule were separated using native PAGE. An 8% resolving gel (4.7 ml H₂O 2.7 ml acrylamide:bis-acrylamide (30%:0.8%), 2.5 ml 1.5M Tris-HCl (pH 8.8), 0.1 ml 10% Ammonium persulfate and 8 µl TEMED made up to 10 ml) A 5% stacking gel (3.1 ml H₂O, 0.67 ml acrylamide:bis-acrylamide (30%:0.8%), 0.5 ml 1.0M Tris-HCl (pH 6.8), 0.04 ml 10% Ammonium persulfate and 4 µl TEMED made up to 4 ml). The running buffer consisted of 25mM Tris-HCl (pH 8.3) and 192mM glycine. The proteins were separated at 0°C and following electrophoresis the gel was equilibrated for 1 hour in 25 mM sodium phosphate buffer (pH 7.4), 1mM CaCl₂ and incubated overnight in 25mM sodium phosphate buffer (pH 7.4), 2% [w/v] UDP-glucose, 1 mM CaCl₂ at room temperature. A lane from the gel was cut off and incubated separately overnight as a control in 25mM sodium phosphate buffer (pH 7.4) and 1 mM CaCl₂ at room temperature. The gels were fixed in a solution of 40% methanol 7% acetic acid for 30 min and oxidized in a solution of 1% periodic acid and 3% acetic acid for 60 min. The gel was washed in ddH₂O to remove the acid. The gel was incubated for 60 min in Schiff's stain (SIGMA-ALDRICH, #S5133-500ml) to visualize the glucan bands.

3.2.2.3 Partial purification of paramylon synthase through product entrapment

Partially purified proteins (approximately 2000 mg), solubilized with CHAPS from the paramylon granule, and were further purified by incubating the protein extract in 20 mM Tris-HCl (pH 7.4), 1 mM CaCl₂, 5 mM UDP-glucose and 20% (m/v) paramylon (Sigma) for 30 min at 25°C followed by 90 min at 4°C with gentle agitation. The proteins associated to the paramylon granule were centrifuged

at 15000 rpm at 4°C for 15 min and the pellet was washed twice in buffer consisting of in 20 mM Tris-HCl (pH 7.4), 1 mM CaCl and 5 mM UDP-glucose. The paramylon granule with the entrapped proteins were incubated overnight at 4° C in 5X Laemmli loading buffer (5ml glycerol, 1g SDS, 2.56 ml β-mercaptoethanol, 2.13ml 0.5M Tris-HCl pH 6.8 and 5mg bromophenol blue). The denatured supernatant was removed for further purification by PAGE.

3.2.2.4 MOPS PAGE of the product entrapped protein

Partially purified proteins were separated at higher resolution using MOPS PAGE gels. Resolving gels were cast in 7 x 11 cm by 1.0 mm plates using a BioRad Protean minigel apparatus (BioRad Laboratories GmbH, Munich, Germany). Resolving gels (10%) consisted of 2.85ml 3.5X gel buffer (1.25 M bis-Tris-HCl (Bis(2-hydroxyethyl) aminotris (hydroxymethyl) methane, Sigma) pH 6.8), 3.33ml acrylamide:bis-acrylamide (30%:0.8%), 3.82ml H₂O, 100μl 10% APS (Ammonium persulfate) and 10μl TEMED. Stacking gels (4%) consisted 2.85ml 3.5X gel buffer, 1.7ml acrylamide:bis-acrylamide (30%:0.8%)(Sigma-Aldrich), 5.45ml H₂O, 100μl 10% APS and 10μl TEMED. A 5X high-MW running buffer (250mM MOPS, 250mM Tris, 5mM EDTA, 0.5% SDS and 5mM NaHSO₃ (sodium bisulfite)) was used for the electrophoresis. Electrophoresis was performed at room temperature at 150V.

Desalted partially purified protein samples were prepared in 5X Laemmli loading buffer (5ml glycerol, 1g SDS, 2.56 ml β-mercaptoethanol, 2.13ml 0.5M Tris-HCl pH 6.8 and 5mg bromophenol blue). Product entrapment samples were suspended and diluted in Laemmli loading buffer. Samples were incubated in loading buffer at 95°C for 5 min. A pre stained PAGE Marker™ Prestained Protein Ladder (Fermentas #SM0671) was used as the molecular marker.

3.2.2.5 Staining of PAGE gels

Gels were fixed in a buffer (25% isopropanol (v/v) and 10% acetic acid (v/v)) for one hour and then rinsed in distilled water. The Colloidal Coomassie Blue Staining Kit (Invitrogen) was used to stain for one hour, and then destained with several washes of distilled water on a rotational shaker.

3.2.2.6 Western blots

Desalted partially purified protein and protein purified by product entrapment was blotted onto Nitrocellulose and Immobulin - PVDF from SDS PAGE and MOPS PAGE gels according to Sambrooke *et al.* (1989). Protein samples were prepared and denatured as described for PAGE gels and a PAGE Ruler™ Prestained Protein Ladder (Fermentas #SM 0671) was used as the molecular marker. Following electrophoresis, PAGE gels were pre incubated in transfer buffer (48mM Tris-HCl, 39mM glycine, 20% methanol and 0.0375% SDS). Separated proteins were transferred from PAGE gels onto membranes with the use of a Trans-Blot SD semi-dry electrophoretic transfer cell (BioRad Laboratories GmbH, Munich, Germany). Membranes were washed in TBST (20mM Tris-HCl pH 7.5, 150mM NaCl and 0.05% Tween 20) and blocked in blocking solution consisting of TBST and 1% BSA (Bovine Albumen (Fraction V), Roche) for two hours. An anti paramylon rabbit raised polyclonal primary antibody (University of Stellenbosch, South Africa), raised against proteins associated to the paramylon granule, was used as the primary antibody. The membranes were washed five times in TBST and incubated in 1% BSA and TBST solution with the secondary antibody in a 1:10000 dilution for one hour. The membranes were subsequently washed 5 times in TBST and rinsed with distilled water and incubated in a solution consisting of one NBT/BCIP Ready-to-use tablet (Roche) in 10ml distilled water, until colour development was observed. The membranes were washed in water to stop the colour development reaction.

3.2.3 Tryptic digestion and LCMSMS of partially purified paramylon synthase

3.2.3.1 Trypsin digestion of excised gel fragments

Following SDS-PAGE gel electrophoresis, protein bands stained with Page Blue™ (Fermentas) were excised from the gel according to Shevchenko *et al.*, 1996. Gel fragments were excised and cut up into smaller fragments and dehydrated at room temperature for 10 min in acetonitrile. The fragments were dried in a speed-vac for 10 min and soaked in 150µl of 10mM DTT and 100mM NH₄HCO₃ for 1 h at 56°C. Gel fragments were subsequently soaked in 150µl of 100mM NH₄HCO₃ and 55mM iodoacetamide for 45min in the dark at room temperature and then washed in 150µl of 100mM NH₄HCO₃ for 10 min at RT. The gel fragments were dehydrated in acetonitrile again for 10 min at RT and washed in 100mM NH₄HCO₃ before being dried for 10min in a speed-vac.

Gel fragments were incubated on ice in 35µl digestion buffer (12.5ng/µl trypsin (Promega sequence-grade modified porcine trypsin, #V511A) and 50mM NH₄HCO₃). The digestion buffer was removed

and the fragments were hydrated in 10µl of 50mM NH₄HCO₃ and cleaved overnight at 37°C. Fragments were centrifuged for 1 min at 13,000 rpm and the supernatant was removed and 20µl of 20mM NH₄HCO₃ was added to the gel fragments and incubated for 10min at room temperature. The removed supernatant was added to the digested gel fragments with 25µl of 5% formic acid and 50% acetonitrile solution and incubated for 10 min at room temperature. The 5% formic acid and 50% acetonitrile was removed by centrifuging the tube for 1 min at 10,000g and the supernatant was saved and dried using a speed-vac.

3.2.3.2 LCMSMS

The digested peptides were sequenced at the Central Analytical Facility (University of Stellenbosch, South Africa). A CapLC coupled to a Waters API Q-TOF Ultima (Massachusetts, USA) was used to sequence the digested peptides which were separated on a capillary LC coupled to the nano-spray (ESI) source of the tandem mass spectrophotometer (QTOF). The probable sequence identity of the peptides was determined using sequence information of collisionally induced dissociation of selected peptide ions against the database.

3.3 Transcriptome sequencing

The *E. gracilis* transcriptome was sequenced using 4µg highly purified cDNA (Section 3.1.1.1 – 3.1.1.4) with the genome sequencer FLX system (Roche 454, USA) at inqaba biotech, (Pretoria, South Africa). The *E. gracilis* genome data was obtained from the MC Field laboratory sequence database (http://web.me.com/mfield/Euglena_gracilis).

Results and Discussion

4.1.1 Construction, amplification, infection and expression of the cDNA library

Euglena gracilis liquid cultures were grown and harvested during exponential growth phase and paramylon synthesis after supplementation of fresh media containing glucose. Total RNA was extracted from harvested cells, quantified and tested for quality as described. The extracted RNA was further purified to obtain high quality mRNA that would express gene products and exclude rRNA. The mRNA was reverse transcribed into cDNA and separated by agarose gel electrophoresis. A gel

fragment of the resulting cDNA smear was excised that included cDNA fragments larger than 1000bp. These cDNA fragments were ligated into the pBluescript based library and packaged into empty phage particles. Phage particles were incubated with XL1-Blue MRF' *E. coli* cells to form a primary library. The primary library was then amplified to make an amplified library by incubating the primary library with XL1-Blue MRF' cells and plating them out, incubating and lysing them from which phages were washed and collected. The amplified cDNA phage library was then used for further screening. A mass excision library was constructed to generate a pBluescript based plasmid library. The library was grown from phage infected SOLAR cells an approximately 1ml at 1µg/µl of the mass excision library was purified.

4.1.2 Screening of the library

4.1.2.1 Screening of the phage library using an antibody

Nitrocellulose membranes were lifted from phage infected bacterial plates where proteins expressed by the phage library were bound. These membranes were screened using a polyclonal antibody raised against the proteins associated to paramylon granular membrane as the primary antibody. The membranes incubated on the plaques were marked for orientation and then blocked. The primary antibody was incubated in blocking solution as described and incubated with a secondary antibody conjugated with an alkaline phosphatase. The membranes were screened for alkaline phosphatase activity to determine the position of plaques expressing proteins that bound to the nitrocellulose membranes. In total, an estimated 186 000 plaques were screened on 116 membranes. Plaques that generated signals and that gave significantly darker signals than neighbour plaques after immunoblotting were identified and isolated through single clone excision, and used to directly infect secondary screens. However, secondary screens never produced a concomitant increase in the number of darker signals. Plaques that did not generate a darker signal were used as negative controls for secondary screens. Variations of exposure with the primary antibody or NBT/BCIP did not produce different results.

The primary antibody, to an extent non-specifically recognized all plaques that formed. It was assumed that despite this, paramylon granule related proteins expressed would have produced darker signals than the non specific plaques. As the polyclonal primary antibody was raised in rabbits, anti *E. coli* antibodies would also be present in the primary antibody serum. During cell lysis, bacterial proteins would also have been bound to the membranes producing a background signal. If

paramylon synthase related proteins were not expressed significantly to generate a darker signal, any positive signals would have been marginalized.

As a significant number of plaques were screened, it can be speculated that problems would have been caused by the non specific recognition of the antibody or possibly due to the expression of the membrane proteins which have different folding conformations and epitopes within the bacterial system. Later experiments confirmed the polyclonal antibody's specificity for solubilized paramylon synthase proteins after isolation from the membranes.

4.1.2.2 Screening of the library using radio-labeled substrate

Nitrocellulose membranes were lifted from phage infected plates where proteins expressed by the phage library were bound. These membranes were screened using the radio-labeled [^{14}C] UDP-glucose for substrate (Bäumer *et al.*, 2001; van der Merwe, 2007). Approximately 33,000 plaques were screened in this way. Protein products expressed during phage lysis were blotted onto the membranes. The blocked membranes were incubated with limited [^{14}C] UDP-glucose and the membranes were stored with phosphor imager screens for up to seven days. Different blocking solutions as well as exposure times were used but did not have significant differences. Images were obtained from the membranes and darker signals, potentially positive clones caused by [^{14}C] UDP-glucose accumulation, which appeared on the membranes were used to identify the corresponding plaques on the plate. These phages were excised, stored and used to directly inoculate secondary screening experiments. Plaques that did not give signal were used with secondary screens as a control. However, secondary screens often did not yield a greater number of positive clones the primary screens and membranes with negative controls often yielded a comparative amount of false positive signals. During secondary screens, membranes that appeared to have a considerably greater number of positive clones relative to the other membranes were then compared to the plaques on the plates and the phages were isolated through single clone excision. During one secondary screen, one membrane showed a considerably greater number of signals compared to the other membranes in the secondary screen and the negative control. Corresponding plaques were purified through single clone excision and the pBluescript vectors that were purified were sequenced. If the signals were caused by a peptide due to affinity for [^{14}C] UDP-glucose it would be expected that the clones excised during the secondary screen would all have similar sequences. However, the sequence data was analysed using NCBI BLAST where none of the excised clones shared sequence similarity to glycosyl transferases or were similar to each other (Table 6). The presence of specific *E. gracilis*

genes, such as tubulin, elongation factors and cytoskeletal proteins, indicate that the library was at least amplified but that the phages were expressing non specific gene products.

Table 6: BLASTx results of false positive clones of the [¹⁴C] UDP-glucose radio-labeled phage screens

Clone nr.	Description of BLASTx hit result	E-value	NCBI Accession nr.
1	60S ribosomal protein L2 [Trypanosoma cruzi]	2.0E-71	ref XP_816366.1
2	cytoskeletal protein [Euglena gracilis]	3.0E-13	gb AAB23241.1
8	receptor of activated kinase C 1 [Chlamydomonas reinhardtii]	2.0E-14	ref XP_001698065.1
9	beta-tubulin [Euglena gracilis]	2.0E-65	gb AAK37834.1 AF182558_1
10	elongation factor 1 alpha [Euglena gracilis]	5.0E-43	gb ACO50110.1

Due to the amount of false positive signals yielded as well as the difference in sequence data from clones that gave signals in secondary screens, it seems that the signals observed were not caused by peptides with affinity for [¹⁴C] UDP-glucose but rather due to non-specific binding or possibly the formation of [¹⁴C] UDP-glucose artifacts on the membranes. Incubation with insufficient amount of [¹⁴C] UDP-glucose to provide a signal, or due to too low expression of the protein may also explain the lack of positive signals. Changes in enzyme conformation due to expression in the bacterial system may also influence the substrate binding affinity of the membrane-bound paramylon synthase enzymes.

4.1.2.3 Functional screening of the mass excision library using aniline blue

Functional screening was performed based on the affinity of aniline blue for β -1,3-glucan polymers. Aniline blue stains β -1,3-glucans yellow under ultraviolet light where colonies transformed with the mass excision library, containing cDNA inserts that express β -1,3-glycosyl transferases, would be expected to stain positively for the product. However, problems with this system is that it relies on expression of active β -1,3-glycosyl transferase, that regulatory peptides are not essential for minimal protein expression, or that the membrane bound paramylon synthase subunits would be functionally expressed in the bacterial cytosol, or that the stain would be able enter the bacterial cell. Some of these problems would be overcome by lysing the cells with lysozyme but this would also create problems during the wash steps. No colonies stained yellow under ultraviolet light and it was assumed that if any paramylon synthase genes were expressed they were not functional in the bacterial system.

4.1.2.4 Colony library screens

Similarly to the functional screen, the mass excision library containing cDNA inserts from *E. gracilis* was transformed into the SOLAR *E. coli* strain. As the library was pBluescript based, the transformed bacterial colonies would express any inserted cDNA. Colonies were blotted onto nitrocellulose membranes, replicated, fixed and then blocked with a blocking solution that included lysozyme. The membranes were not washed after lysis although fresh blocking buffer was added during incubation with the primary antibody. It was thought that the remaining lysate on the membranes and in the Petri dishes would be enough to reduce non specific binding of *E. coli* proteins. However, non specific signals still appeared in the screens. Secondary screens were conducted, using colonies from replica membranes, which gave considerably darker signals. Colonies that did not give darker signals were used as negative controls. None of the colonies during the second rounds of screening gave significantly darker signals relative to the control.

4.1.3 Screening of cDNA using PCR

Isolated cDNA from *E. gracilis* was used as template for PCR reactions using degenerate and non degenerate primers designed from a partially purified peptide, with UDP-glucose binding affinity, through SDS-PAGE and LCMS, from protein partially purified and solubilized from the paramylon granular membrane (van der Merwe, 2007). Primers were redesigned based on that peptide sequence but with a 4 fold degree of degeneracy for this study. The peptide sequence from a 37 kDa peptide, which had UDP-glucose binding affinity, showed identity to the fungal glucan synthases of *Debaryomyces hanseii* and KRE6 β -1,6-glucan synthase of *Candida albicans* and *Saccharomyces cerevisiae*. The newly designed primer sequences included conserved WN amino acids on the N-terminal which were derived from the *Debaryomyces hanseii*, which increased the length of the primer without increasing the degeneracy. The newly designed degenerate and non-degenerate primers were used to screen purified cDNA using PCR. Conserved domains derived from β -1,3-glucan synthases of several organisms were also used to design degenerate primers (Table 4). These primers were used with oligo(dT) primers to screen *E. gracilis* cDNA. Several bands were amplified and excised and purified from agarose gels, cloned and sequenced. When the sequences were screened using NCBI BLASTx, no glycosyl transferase related sequences were identified.

The mass excision library was screened using the conserved domains primers (CD) (Table 3). As the mass excision library was designed in pBluescript, the inserted cDNA is flanked by both T3 and T7 primers. Due to the directional insertion into the vector (Zap cDNA Synthesis kit protocol), the T3

primer would be able to be paired with a gene specific reverse primer. The conserved, degenerate and non-degenerate reverse primers were paired with the T3 primer and the forward primers were paired with the T7 primer which were used to amplify DNA from the mass excision library. Several bands per PCR, due to the nature of the cDNA library, were amplified, cloned, purified and sequenced. Sequence data (Table 7) from the amplified bands of the mass excision library did not show identity to known glycosyl transferases.

Table 7: PCR results of cDNA and the mass excision library amplified and sequenced.

Primer pair	BLASTx description of the sequence	Accession number	E-value
cDNA template			
rev CD1 - oligo(dT)	aldehyde dehydrogenase (acceptor) [<i>Trichodesmium erythraeum</i> IMS101]	ref YP_722268.1	9.0E-43
rev CD2 - oligo(dT)	hypothetical protein LOC100187597 [<i>Nasonia vitripennis</i>]	ref NP_001127800.1	3.1E+00
rev CD3 - oligo(dT)	Receptor of activated protein kinase C 1A, [<i>Ostreococcus lucimarinus</i>]	ref XP_001421096.1	6.0E-20
rev CD4 - oligo(dT)	PREDICTED: similar to ribosomal protein L8 [<i>Strongylocentrotus purpuratus</i>]	ref XP_796001.1	3.0E-55
rev CD5 - oligo(dT)	beta-tubulin [<i>Euglena gracilis</i>]	gb AAK37834.1 AF182558_1	4.0E-73
fwd non deg - oligo (dT)	ubiquitin-protein ligase, putative [<i>Trypanosoma brucei gambiense</i>]	emb CBH13171.1	1.8E+00
fwd non deg - oligo (dT)	putative ribosomal protein L15 [<i>Diaphorina citri</i>]	gb ABG81972.1	1.0E-35
rev deg - oligo (dT)	Adenosylcobalamin-dependent ribonucleoside-triphosphate reductase [<i>Euglena gracilis</i>]	sp Q2PDF6.1 RTPR_EUGGR	6.0E-45
rev deg - oligo (dT)	PREDICTED: thrombospondin, type I, domain containing 7B isoform 2 [<i>Oryctolagus cuniculus</i>]	ref XP_002712172.1	1.3E+00
fwd deg - oligo (dT)	ADP-ribosylation factor 1 [<i>Trypanosoma congolense</i>]	gb ABT17155.1	4.0E-44
Mass excision library template			
rev CD1 - T3	cytoskeletal protein [<i>Euglena gracilis</i>]	emb CAA78364.1	9.0E-13
rev CD2 - T3	ATP-dependent Clp protease ATPase subunit [<i>Geobacillus kaustophilus</i> HTA426]	ref YP_145931.1	5.5E+00
rev CD3 - T3	similar to activated kinase C receptor [<i>Trypanosoma carassii</i>]	gb ABS01292.1	5.0E-17
rev CD4 - T3	5-methyltetrahydropteroyltriglutamate/homocysteine S-methyltransferase [<i>R. pickettii</i>]	ref YP_001893059.1	2.0E-22
rev CD5 - T3	beta-tubulin [<i>Euglena gracilis</i>]	gb AAK37834.1 AF182558_1	3.0E-72
fwd CD6 - rev CD7	elongation factor 1 alpha [<i>Euglena gracilis</i>]	gb ACO50110.1	4.0E-74
fwd CD7 - rev CD2	60S ribosomal protein L12 [<i>Sclerotinia sclerotiorum</i> 1980]	ref XP_001586492.1	2.0E-08
fwd CD2 - rev CD1	5-carboxymethyl-2-hydroxymuconate Delta-isomerase [<i>Geobacter</i> sp. M21]	ref YP_003020495.1	2.0E-12
fwd CD1 - rev CD3	predicted protein [<i>Thalassiosira pseudonana</i> CCMP1335]	ref XP_002294414.1	7.0E+00
fwd CD4 - rev CD5	predicted protein [<i>Physcomitrella patens</i> subsp. patens]	ref XP_001772298.1	1.0E-05
T3 - rev CD1	spore coat protein (outer) [<i>Bacillus subtilis</i> subsp. subtilis]	ref NP_391486.1	8.3E-02
T3 - rev CD2	heat shock protein 60 [<i>Euglena gracilis</i>]	emb CAA65238.1	8.0E-41
T3 - rev CD3	plasma membrane H ⁺ -ATPase [<i>Plantago major</i>]	emb CAL35828.1	5.0E-04
T3 - rev CD4	ribosomal protein L13A [<i>Schizophyllum commune</i> H4-8]	ref XP_003038672.1	3.0E-14
T3 - rev CD6	ribulose-bisphosphate carboxylase [<i>Euglena gracilis</i>]	emb CAA55779.1	5.0E-99
T3 - rev CD7	6-phosphogluconolactonase [<i>Chryseobacterium gleum</i> ATCC 35910]	ref ZP_07087499.1	6.9E+00
T7 - fwd CD1	spore coat protein (outer) [<i>Bacillus subtilis</i> subsp. subtilis]	ref NP_391486.1	1.4E-01
T7 - fwd CD2	eukaryotic initiation factor 4E, putative [<i>Phytophthora infestans</i> T30-4]	ref XP_002907026.1	2.0E-23
T7 - fwd CD3	vacuolar ATP synthase [<i>Leishmania infantum</i> JPCM5]	ref XP_001465476.1	2.0E-31
T7 - fwd CD4	succinyl-CoA synthetase subunit beta [<i>Rickettsia typhi</i> str. Wilmington]	ref YP_067379.1	4.0E-04
T7 - fwd CD7	hypothetical protein [<i>Trichomonas vaginalis</i> G3]	ref XP_001322505.1	4.0E-04

4.2 Protein purification

4.2.1 Paramylon synthase crude protein extraction

Euglena gracilis cells, grown under heterotrophic conditions, were centrifuged and disrupted by sonication on ice in buffer containing protease inhibitors and DTT. After disruption the paramylon granules were separated and washed of the cellular membranes and cell debris by centrifugation. After the paramylon granules were washed, the paramylon synthases and membrane associated proteins were solubilized from the granule membranes with the detergent CHAPS. Activity was determined by Schiff staining native PAGE. Previous studies have shown paramylon synthase activities from crude protein extracts using UDP-glucose assays, Schiff staining and Coomassie colloidal blue staining (Bäumer *et al.*, 2001; van der Merwe, 2007).

4.2.2 FPLC, PAGE separation, product entrapment and immunoblotting of partially purified paramylon synthase proteins

4.2.2.1 FPLC of paramylon synthase proteins

The polyclonal antibody was purified using an ACTA prime FPLC and a HiTRAP Protein A column. Elution of the polyclonal antibody from the column successfully separated the polyclonal antibody from serum contaminants and proteins (Figure 4). The purified polyclonal antibody was coupled to a CNBr-activated sepharose column and partially purified protein extract was run through the column and allowed to bind to the antibody based on specific affinity for the polyclonal antibody. The resultant peaks indicated that protein that had affinity for the coupled antibody was eluted when the binding buffer was replaced with the elution buffer (Figure 5). The first peak represented all other non-specific proteins without antibody affinity while the second minor peak represented the protein with antibody affinity. As the total protein loaded was ~5mg the relative peak sizes suggest a very low concentration of proteins with antibody binding specificity would have been eluted which was not detectable in subsequent ammonium sulfate precipitation purification steps.

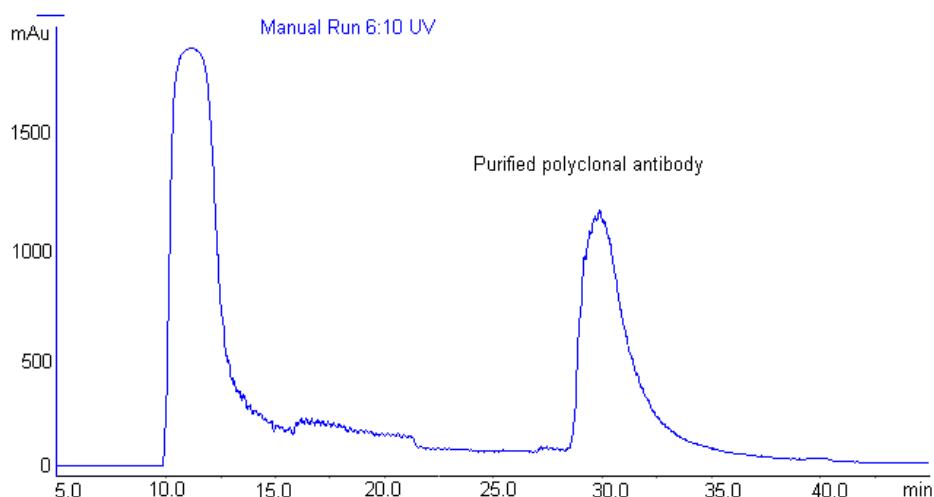


Figure 4: FPLC purification of polyclonal anti-paramylon synthase antibodies. The antibodies were eluted at 30 min before being coupled to CNBr-activated sepharose.

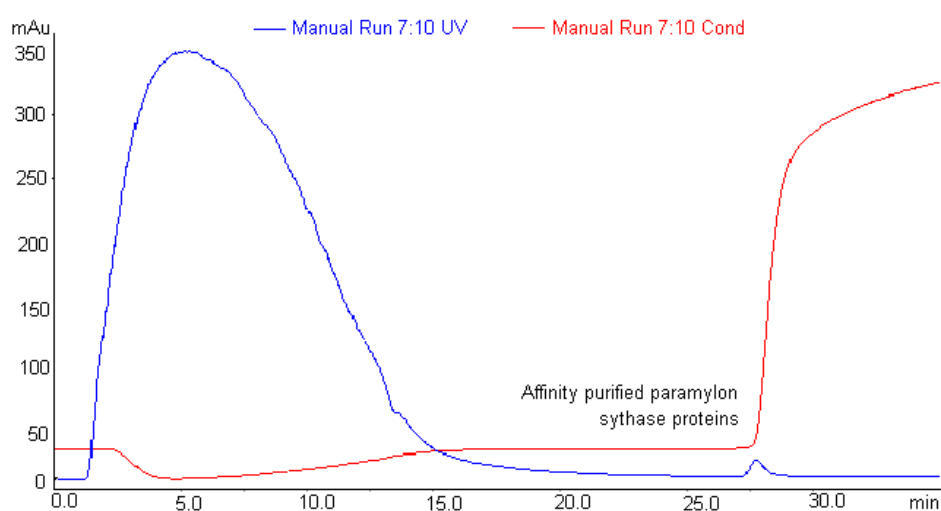


Figure 5: FPLC purified paramylon synthase proteins were purified through the polyclonal antibody bound CNBr-sepharose column after the binding buffer was exchanged with the elution buffer after non specific proteins were washed out.

4.2.2.2 Detection of in gel paramylon synthase activity

Paramylon synthase activity was detected in protein solubilized from the granule membranes by separating solubilized undenatured protein on native PAGE followed by staining the gels in Schiff stain solution. Solubilized glucan synthases are known to remain stable and active in gels (Thelen and

Delmer, 1986). Paramylon as a primer was not incubated in the reaction with the protein extract although Ca^{2+} cation cofactors and UDP-glucose as substrate were added during incubation.

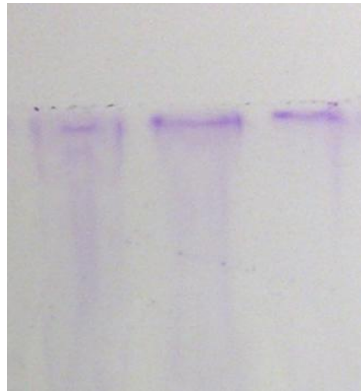


Figure 6: Schiff stained Native PAGE of in-gel paramylon synthase activity. Partially purified CHAPS solubilized protein was loaded undenatured into the three lanes. Schiff staining of the gel showed a large presumably ~670 kDa protein complex (Bäumer et al, 2001). Control gel did not stain a product (Data not shown).

Schiff staining of the native PAGE showed product formation by a high molecular weight protein complex migrating only slightly into the resolving gel (Figure 6) but not in the control (Data not shown). As this corresponds with previous results (van der Merwe, 2007) and as *E. gracilis* does not produce other carbohydrate polymer products such as cellulose, glycogen or starch the product was taken to be paramylon. The paramylon synthase complex was previously determined to be 670 kDa using a similar method (Bäumer *et al.*, 2001).

4.2.2.3 Paramylon synthase partial purification through product entrapment

CHAPS solubilized crude protein extract was dialyzed to further purify and concentrate the protein extract. The partially purified protein was quantified. A method of protein purification, specifically for β -glucan synthases has been through product entrapment (Kang *et al.*, 1984) where many of the β -1,3-glucan synthases from plants and some fungi have been successfully partially purified and enriched using this method. Product entrapment of protein relies on the association of the enzyme complex with the product. Partially purified protein is incubated with the purified product and allowed to associate with the insoluble product. Insoluble purified paramylon (SIGMA) was used as the product which was incubated at room temperature with dialyzed protein extract, Ca^{2+} and UDP-glucose substrate. In literature, several unidentified proteins were non-specifically purified with the

β -1,3-glucan synthase proteins. As several functionally different proteins are involved in the synthesis of paramylon it was expected that several proteins would be enriched with the β -1,3-glucan synthase proteins along with non-specifically enriched proteins. Following incubation, the insoluble protein-paramylon entrapment was separated from the unbound proteins using a filtered spin column. The filter of the spin column with the residual protein-paramylon entrapment was removed and separated using MOPS-PAGE (Figure 7, p 40).

4.2.2.4 MOPS-PAGE of the product entrapped protein

The protein-paramylon entrapment was separated using MOPS-PAGE. The protein-paramylon entrapment was denatured by heat and incubation with SDS loading buffer. The SDS loading buffer diluted the purified protein and the denatured protein was spun down to remove the insoluble paramylon. The MOPS-PAGE separated several protein bands with high resolution, of which several appeared enriched after staining with Coomassie. Although many faint bands were also separated. Previous literature identified the sizes of the β -1,3-glucan synthase catalytic subunits using radio-labeled substrates or immune-precipitation. Many of the faint bands were then excluded as non-specific proteins.

Previous literature (Bäumer *et al.*, 2001) identified two bands, 55 and 37 kDa which bound 32 [P] UDP-glucose, purified with sucrose density gradient ultracentrifugation. Coomassie staining of the same protein fractions showed several unidentified bands separating at approximately 20, 60, 62 and 72 kDa as well as the 55 and 37 kDa bands. These other bands were never investigated, as the proteins did not bind UDP-glucose, although may still be involved in the glucan synthase complex or associated with the membrane as structural or transport proteins. In more recent studies in this lab (van der Merwe, 2007), similar experiments to Bäumer's group identified proteins separating in a similar pattern for Coomassie stained PAGE gels as well as for 32 [P] UDP-glucose in-gel assays, except that a protein of approximately 80 kDa also bound 32 [P] UDP-glucose but was never further analyzed. Although these experiments separated and purified the proteins using different methods, similar patterns were observed with those purified and enriched using product entrapment.

4.2.2.5 Western Blot and gel excision of paramylon synthase proteins

Product entrapped peptides were separated using MOPS-PAGE where separate gels were stained with Coomassie colloidal blue and immunoblotted using a primary antibody raised against

membrane proteins associated to the granule (Figure 7). The western blot identified several proteins separated by product entrapment purification which correlated to proteins enriched on the MOPS-PAGE gel. The proteins that were correlated with the bands on the western blot were excised from the MOPS-PAGE gel and treated with proteases. Seven bands were excised from the MOPS-PAGE gel of approximately 37, 55, 60, 62, 72, 80 and 95 kDa.

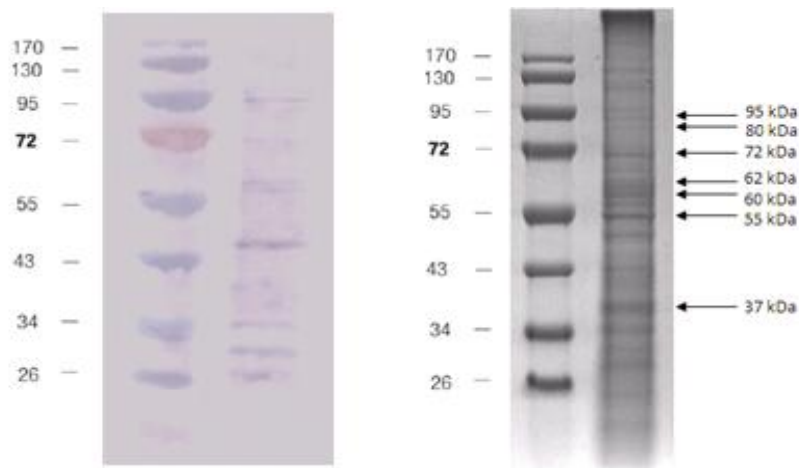


Figure 7: Western Blot (left) analysis of partially purified product entrap proteins associated to the paramylon granule (right lanes). Seven proteins were separated with MOPS-PAGE (right), which corresponded to the western blot, were excised from the gel and digested using trypsin.

4.2.3 LCMSMS

The excised proteins were digested and short peptide sequences were predicted from each of these proteins. The results of the digested peptides were analyzed using the MASCOT program where each of the excised proteins from the partially purified from the paramylon granular membrane was digested, separated and grouped into queries. Each query contained predicted short peptide sequences which were analysed using BLAST and searched against the protein database (BLASTp) of NCBI where the NCBI reference numbers and the specific gene names were recorded in addendum A. For each short peptide, the first 100 BLAST hits were searched and the relevant genes were recorded, specifically β -1,3-glucosyl transferases, glycosyl transferase gene families and calcium and UDP-glucose binding proteins. Of approximately 900 predicted peptide queries, 289 peptides showed identity or partial identity to known glycosyl transferases. As β -1,3-glucan synthases fall under glycosyl transferase family 2 (GT2), with mannosyl transferases, chitin synthases and cellulose synthases and all having common GT-A and GT-B folding domains, predicted query peptides with

similar identities to GT-2 glycosyl transferases were selected. NCBI BLAST results did not usually have complete identity with the query peptides and the query peptides showed similar identity to many non-specific proteins with similar or even higher degrees of identity. E-values were recorded but were relatively very high due to the nature of the short peptide sequences.

4.3 Sequence analysis of the *E. gracilis* transcriptome and genome

4.3.1 Sequencing of the *Euglena gracilis* transcriptome

The *E. gracilis* transcriptome was partially sequenced using the Genome Sequencer FLX system. High quality RNA was extracted from *E. gracilis* liquid cultures during exponential growth and paramylon synthesis. The total RNA was extracted and quantified from which mRNA was further purified by selecting mRNA sequences with poly-(T) tails most likely to express gene products. The purified mRNA was quantified and the quality of the sample was controlled using agarose gel electrophoresis. The mRNA was converted into double stranded cDNA and quantified and qualified (Figure 8). The prescribed amount of cDNA was purified and sent for sequencing.

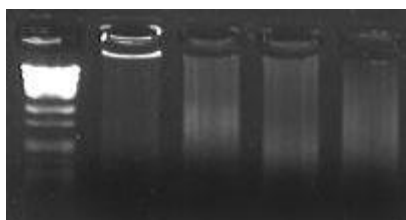


Figure 8: Purified double stranded cDNA smears (lanes 2-5) were pooled and sequenced using the genome sequencer FLX system.

The sequenced cDNA was assembled and arranged into overlapping reads called contigs using CLC genomics workbench. Normally, query transcriptomes are characterized against well characterized reference transcriptomes by directly comparing the generated FASTA sequences. Unfortunately in the case of *Euglena*, no reference sequence libraries are available and only limited sequence information is available.

The approach in this study was to directly sequence the transcriptome and BLAST annotate the available contigs to identify any glycosyl transferases, specifically UDP-glycosyl transferases, β -1,3-glycosyl transferases as well as GT-2 glycosyl transferases. Paramylon synthase genes most likely

originated from genes inherited through the secondary endosymbiotic event (Martin *et al.*, 1993; Archibald *et al.*, 2003; Timmis *et al.*, 2004) as β -1,3-glucans do not appear in any kinetoplastid while they do appear in the forms of callose in plants (Samuels *et al.*, 1995) and as cell wall constituents in fungal systems (Cabib *et al.*, 1987). It is highly likely that paramylon synthase genes would have evolved from cellulose, callose and chitin synthases as opposed to the prokaryote β -1,3-glucans or kinetoplastid glycoproteins.

ORF's for potential prokaryotic and eukaryotic genes were identified from the contigs using the BLASTx function on CLC Genomics Workbench (contig assembly is summarized on Table 8 and Figures 9 and 10), including a pre-existing *Euglena* EST database (tbestdb, <http://tbestdb.bcm.umontreal.ca>), and arranged into contig queries with the reference NCBI accession numbers, number of reads, greatest read length and the lowest E value (data not shown). None of the contigs identified through BLASTx showed any identity to glycosyl transferases or for Ca^{2+} and UDP-glucose binding genes. With such a low number of contigs (1179), which only covered a fraction of the entire transcriptome, no glycosyl transferases were identified although many of the genes identified coded for hypothetical proteins and proteins of unknown function as was expected.

Following from that approach, the short peptide sequences (appendix A) obtained from partially purified proteins associated to the paramylon granule as well as the conserved domains from β -1,3-glucan synthases were used to screen against the partial transcriptome directly. The contigs were converted into ORF's in FASTA format using the GeneMark heuristic model (Besemer and Borodovsky, 1999). Both eukaryote and prokaryote ORF's were converted into protein FASTA sequences and used as a database to screen against using the paramylon synthase short peptide fragments (appendix A). As the peptides were expected to have complete identity with the paramylon synthase proteins, they were used directly to screen the ORF's. However, none of the peptide sequences matched any of the available ORF's.

Table 8: Contig assembly of the transcriptome

Assembled reads		Contig count	
Read count	2960	Contig count	1179
Mean read length	746.52	Mean contig length	702.15
Total read length	2209691	Mean coverage	2

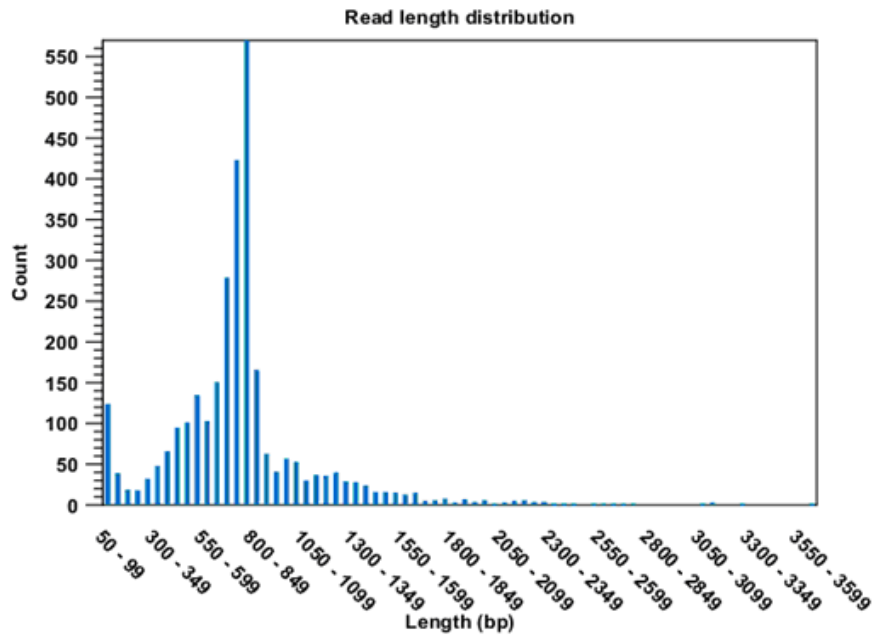


Figure 9: Read length distribution of the sequenced transcriptome.

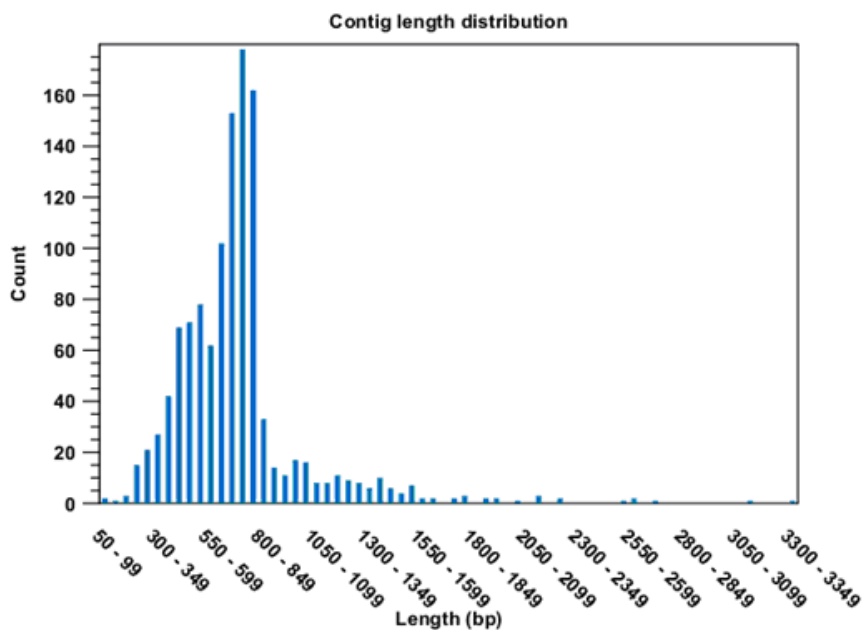


Figure 10: Contig length distribution of the sequenced transcriptome.

The absence of any glycosyl transferases from the transcriptome sequences (data not shown), with the exception of an α -1,6-mannosyl transferase from a different GT-family, where glycosyl transferases are estimated to account for about 1% of universal ORF's (Coutinho *et al.*, 2003) gives an indication of the low read yield. Although a great number of predicted short peptide fragments were identified with identity to glycosyl transferases, they all would be derived from the same proteins.

4.3.2 BLASTing the partially completed *Euglena gracilis* genome sequence

During the course of this study, a *E. gracilis* genome sequencing project was completed and released. The genome sequencing project was aimed at phylogenetic sequencing studies and was released on their website (http://web.me.com/mfield/Euglena_gracilis). The assembled contigs were obtained and BLAST searched using BLASTx and recorded. Initially, the short predicted peptide sequences were used to screen the transcriptome similar to the sequenced transcriptome of this study. Peptide sequences that showed some similarity to the contigs were recorded (Table 9 and 10), particularly where two peptides matched the same contig in the same reading frame. These contigs were BLASTed but did not show identity to any glycosyl transferase sequences. Although two partial sequences NFVNGI and LPPPPRTEL of the query peptide sequences, where the adjacent numbers designate nucleotide position on the contig subject sequences, followed on from each other. As the query LCMSMS peptides are only predicted sequences that were identified as possible glycosyl transferases using BLASTp, there is room for error occurring both with peptide prediction as well as BLAST identification. Many of the query peptides showed some identity with the contigs however most of the contig sequences have relatively short sequence lengths possibly resulting in predicted peptide pairs to be missed. Due to the lack of reference proteins, it may also be possible that the peptide sequences covered variable regions among β -glycosyl transferases and were not identified by BLAST.

Glycosyl transferases are highly variable between organisms and only share or partially share specific motifs and domains as described previously. Without the specific domains being present in the contig queries it is difficult to determine whether the contig is or is not a β -1,3-glycosyl transferase. As the highly conserved GT-2 motifs consist of only three or four amino acids and sometimes as single amino acids, it becomes very difficult to predict, confirm or rule out the contig query sequences that show identity to variable regions of the subject GT-2 glycosyl transferase proteins without further sequencing, cloning and screening for activity. Due to the nature of the assembled transcriptome, many of the contigs are incorrectly assembled having no consensus sequences, repeated and have sequence substitution and frame shift errors. When previously identified *E. gracilis* genes on NCBI were BLAST searched against the *E. gracilis* transcriptome, the genes would be separated over several contigs with minor substitution and deletion sequence errors. Sequence errors causing frame shifts would cause potential positive sequences to be missed using BLASTx.

The sequenced genome consisted of 29613 assembled contigs assembled from ~4 million reads which were BLAST searched using the NCBI BLASTx function of CLC Genomics Workbench. All BLAST results were filtered for “glycosyl transferases”, “glucan”, “cellulose” and “chitin” synthases and recorded (Data not shown). GT-2 sequences with the lowest E-values and highest identities were recorded along with the conserved domains of the subject sequence (Table. 10).

Several potential β -1,3-glucan synthase partial sequences were identified along with GT-2 glycosyl transferase and cellulose synthase sequences using BLASTx. Candidate sequences which partially aligned with GT-2 glycosyl transferases are represented with their corresponding E-values, related conserved domains and motifs (Table 10). Contigs that had significant identity to β -1, 3-glucan synthases and GT-2 glycosyl transferases were aligned (Table 11). Two contigs showed significant partial identity with β -glycosyl transferases; contig 11792 (Figure 13, p 49) and contig 11487 (Figures 14 and 15, p 51). The expect values of the alignments are relatively high and the queries align only partially with the subject sequences.

Table 9: Predicted peptide sequences aligned with the contigs of the sequenced genome

Peptide sequence	Protein	Peptide BLAST description	Contig	E-value	ORF	Contig sequence match
LSSASAPATK	80 kDa	ref XP_001568074.1 GTP-binding elongation factor tu protein [<i>Leishmania braziliensis</i>]	contig0243	9.8	+1	130- SSSASPA -150
GPSSPSSADACR	35 kDa	ref ZP_06276423.1 β -1,3-glucan synthesis-like protein [<i>Streptomyces</i> sp.]	contig0243	4.0	+1	157- PSSPASASGAC -189
MLIPVMRWQR	80 kDa	ref YP_001634862.1 glycosyl transferase group 1 [<i>Chloroflexus aurantiacus</i> J-10-fl]	contig0650	7.3	+1	241- IPVLRW -258
EVRIGSASVR	35 kDa	ref YP_003013567.1 Peptidoglycan glycosyltransferase [<i>Paenibacillus</i> sp. JDR-2]	contig0650	3.0	+1	520- VRIGSANV -543
KYYDDSPWQR	80 kDa	ref ZP_01092700.1 putative glycosylhydrolase [<i>Blastopirellula marina</i> DSM 3645]	contig0102	5.5	-2	3037- DDSPWLR -3017
WDYAVVASGGFFIAGK	50 kDa	ref YP_152483.2 peptidoglycan synthetase [<i>Salmonella enterica</i> subsp. ATCC 9150]	contig0102	9.4	-2	1930- AVVASGG -1910
ISSHSLEELAAELQISKEK	62 kDa	ref YP_001450314.1 glycosyltransferase [<i>Streptococcus gordonii</i> str. Challis substr. CH1]	contig0179	5.1	-3	555- SHSLDPLSAE -526
FTRDWTk	60 kDa	ref YP_001404698.1 glycosyl transferase, group 1 [<i>Candidatus Methanoregula boonei</i>]	contig0179	1.2	-3	462- TRDWTk -445
QVNSIYGKFSASWK	60 kDa	ref YP_003182856.1 glycosyl transferase family 51 [<i>Eggerthella lenta</i> DSM 2243]	contig0075	7.1	+2	548- NSIFQKFSa -574
GTGCSDDLPPPSRAEMGR	50 kDa	ref YP_001537113.1 glycosyl transferase group 1 [<i>Salinispora arenicola</i> CNS-205]	contig0075	6.9	+2	140- SPDLPP -157
GLGMLGWVFFPPREDLSGGIDR	35 kDa	ref ZP_03997417.1 glycosyltransferase [<i>Halogeometricum borinquense</i> DSM 11551]	contig0470	3.7	+3	93- PPREDL -110
SGFFSHCCLSFSPK	50 kDa	ref ZP_05724441.1 glycosyl transferase group 1 [<i>Dickeya dadantii</i> Ech586]	contig0470	4.0	+3	27- FGNCCEFSP -53
		ref ZP_06115144.1 glycosyl transferase, group 1 [<i>Clostridium hathewayi</i> DSM 13479]	contig0210	9.6	-2	228- CCLFFPS -208
GPSSPSSADACR	35 kDa	ref ZP_06276423.1 β -1,3-glucan synthesis-like protein [<i>Streptomyces</i> sp. ACT-1]	contig0210	7.2	-2	543- SPSTSPAACR -514
VKGNFVNLPGDTLEK	50 kDa	ref YP_679052.1 glycosyltransferase-S [<i>Cytophaga hutchinsonii</i> ATCC]	contig2812	7.0	+2	68- NFVNGL -85
		ref ZP_04061534.1 glycosyltransferase-I [<i>Streptococcus salivarius</i> SK126]	contig0852	7.0	-2	168- NFVNGL -151
			contig0684	7.0	+1	574- NFVNGL -591
GTGCSDDLPPPSRAEMGR	50 kDa	ref YP_001537113.1 glycosyl transferase group 1 [<i>Salinispora arenicola</i> CNS-205]	contig2812	6.9	+2	86- LPPPPRTEL -112
			contig0852	6.9	-2	150- LPPPPRTEL -124
			contig0684	6.9	+1	592- LPPPPRTEL -618
			contig1856	6.9	-3	216- DLPPPS -199
GAFHTVSLQKYSK	35 kDa	ref ZP_03395721.1 glycosyl transferase, group 2 family protein [<i>Pseudomonas syringae</i>]	contig1856	7.2	-3	21- LQKYSK -4

Table 10: BLASTx results of the genome contigs aligned with two predicted peptide sequences.

Contig	Contig BLASTx description	E-value
contig02437	no hits for specific reading frame	
contig06506	no hits for specific reading frame	
contig01027	no hits for specific reading frame	
contig01797	XP_002999687.1 predicted protein [<i>Verticillium albo-atrum</i> VaMs.102]	5.7
contig00754	no hits for specific reading frame	
contig04704	YP_890301.1 DNA repair protein Rada [<i>Mycobacterium smegmatis</i> str. MC2 155]	3.4
contig02102	no hits for specific reading frame	
contig28129	XP_002564567.1 Pc22g05320 [<i>Penicillium chrysogenum</i> Wisconsin 54-1255]	3.4
contig08522	XP_002564567.1 Pc22g05320 [<i>Penicillium chrysogenum</i> Wisconsin 54-1255]	3.4
contig06849	no hits for specific reading frame	
contig18568	no hits for specific reading frame	

Table 11: Contig sequences from the sequenced genome with partial identity to GT-2 glycosyl transferase protein sequences.

Contig	BLASTx description of similar proteins and their accession numbers	E-value	Domain	Motif	Domain descriptions of the subject sequence
Contig 2502	YP_002975125.1 cellulose synthase subunit B [<i>Rhizobium leguminosarum</i>]	3.00E+00	pfam03170 PRK11114		BcsB; Bacterial cellulose synthase subunit cellulose synthase regulator protein
Contig 2607	YP_480685.1 glycosyl transferase family protein [<i>Frankia sp.</i> Ccl3]	3.00E+00	cl11394		Glyco_tranf_GTA_type; Glycosyltransferase family A (GT-A)
Contig 3502	XP_002793109.1 1,3-beta-glucanosyltransferase gel2 [<i>Paracoccidioides brasiliensis</i> Pb01]	9.20E+00	cl12144		glycosyl hydrolase family 5
Contig 5589	CBJ30317.1 Cellulose synthase (UDP-forming), family GT2 [<i>Ectocarpus siliculosus</i>]	4.00E+00	cd06421		CESA_CelA like cellulose synthase GT-2
Contig 9141	YP_001682980.1 glycosyl transferase family protein [<i>Caulobacter sp.</i> K31]	7.00E-02	cd00761	DXD	Glyco_tranf_GTA_type; Glycosyltransferase family A (GT-A)
Contig 10917	ZP_01857650.1 glycosyl transferase, group 2 family [<i>Planctomyces maris</i> DSM 8797]	2.00E+00	COG0463		Glycosyltransferases involved in cell wall biogenesis
Contig 11487	XP_001258384.1 1,3-beta-glucanosyltransferase, putative [<i>Neosartorya fischeri</i> NRRL 181]	4.00E+00	cl06842		glycosyl hydrolase family 17
	EDP49809.1 1,3-beta-glucanosyltransferase, putative [<i>Aspergillus fumigatus</i> A1163]	8.00E+00	cl12144		glycosyl hydrolase family 5
Contig 11792	XP_002999868.1 cellulose synthase catalytic subunit [<i>Verticillium albo-atrum</i> VaMs.102]	1.00E+01	cd06421 PRK11204	DXD QXXRW	CESA_CelA like cellulose synthase GT-2 N-glycosyl transferase
	XP_001259116.1 glycosyl transferase, putative [<i>Neosartorya fischeri</i> NRRL 181]	1.00E+01	cd06421	DXD	CESA_CelA like cellulose synthase GT-2
	XP_001273558.1 glycosyl transferase, putative [<i>Aspergillus clavatus</i> NRRL 1]	1.00E+01	cd06421	DXD	CESA_CelA like cellulose synthase GT-2
	XP_748682.1 glycosyl transferase [<i>Aspergillus fumigatus</i> Af293]	1.00E+01	cd06421	DXD	CESA_CelA like cellulose synthase GT-2
Contig 12312	YP_003695065.1 Cellulose synthase BcsB [<i>Starkeya novella</i> DSM 506]	6.00E-01	PRK11114		cellulose synthase regulator protein
Contig 12738	YP_003637451.1 glycosyl transferase family 2 [<i>Cellulomonas flavigena</i> DSM 20109]	7.00E-01	cd04186		GT_2_like_c; Subfamily of GT2 (GT-A) of unknown function.
Contig 14629	CBJ28249.1 Cellulose synthase (UDP-forming), family GT2 [<i>Ectocarpus siliculosus</i>]	2.00E+00	cd06421	DXD	CESA_CelA like cellulose synthase GT-2
Contig 15267	EFL91633.1 glycosyltransferase II [<i>Candidatus Regiella insecticola</i> LSR1]	9.00E-01	cd11394		Glyco_tranf_GTA_type; Glycosyltransferase family A (GT-A)
Contig 16157	YP_003411251.1 glycosyl transferase family 2 [<i>Geodermatophilus obscurus</i> DSM 43160]	2.00E-01	cd04186		GT_2_like_c; Subfamily of GT2 (GT-A) of unknown function.
Contig 17137	ZP_06766496.1 glycosyltransferase, group 2 family [<i>Bacteroides xylanisolvens</i>]	1.00E+00	cd00761		Glyco_tranf_GTA_type; Glycosyltransferase family A (GT-A)
Contig 17339	YP_001419299.1 putative beta (1-6) glucans synthase [<i>Xanthobacter autotrophicus</i> Py2]	3.20E+00	cl02191		Glyco_hydro_17; Glycosyl hydrolases family 17
Contig 17999	YP_001112741.1 glycosyl transferase family protein [<i>Desulfotomaculum reducens</i> MI-1]	2.00E+00	cd02511		UDP-glucose beta 1,4 Glycosyltransferase
Contig 19222	YP_003130477.1 glycosyl transferase family 2 [<i>Halorhabdus utahensis</i> DSM 12940]	9.10E+00	cd06423		CESA_like; CESA_like is the cellulose synthase superfamily
Contig 23477	YP_003576339.1 family 2 glycosyl transferase [<i>Rhodobacter capsulatus</i> SB 1003]	3.70E-01	cd11394		Glyco_tranf_GTA_type; Glycosyltransferase family A (GT-A)
Contig 24032	YP_003072318.1 glycosyltransferase, group 2 family [<i>Teredinibacter turnerae</i> T7901]	7.00E+00	cd06420		GT2 GalNAc transferase
Contig 27271	YP_003239098.1 glycosyl transferase family 2 [<i>Ammonifex degensii</i> KC4]	8.00E-01	cd04186 cd11394		GT_2_like_c; Subfamily of GT2 (GT-A) of unknown function. Glyco_tranf_GTA_type; Glycosyltransferase family A (GT-A)

Contig 3502

|XP_002793109.1| 1,3-β-glucanosyltransferase gel2 [*Paracoccidioides brasiliensis* Pb01]

Length=471, Score = 33.5 bits (75), Expect = 9.2, Identities = 16/53 (30%), Positives = 28/53 (52%), Gaps = 0/53 (0%) Frame = +2

```
Query   233   VQSYFLNIPRIFHLNVQLKRSMYISFFNGSGSFTATAIVTPAAHRLLPRTESA   391
        +Q  +N  RI++LN      + +S FN +G +      + +P AH+ L R + A
Sbjct   77    LQRLGVNTIRIYNLNPATNHDLCVSIFNAAGIYLLIDVNSPIAHQSLNRADPA   129
```

Contig 10917

|ZP_01857650.1| glycosyl transferase, group 2 family protein [*Planctomyces maris* DSM 8797]

Length=304, Score = 37.0 bits (84), Expect = 2E+00, Identities = 15/26 (58%), Positives = 22/26 (85%), Gaps = 1/26 (4%), Frame = +3

```
Query   27    ILGQFWAKKTR-KPSKIFFAQNHDDH   101
        +L  +FWA+K R +PSKIFF++ +DHD
Sbjct   271   VLARFWAQKLRVRPSKIFFSKNDHDD   296
```

Contig 11487

|XP_001258384.1| 1,3-beta-glucanosyltransferase, putative [*Neosartorya fischeri* NRRL 181]

Length=540, Score = 38.5 bits (88), Expect = 4E00 Identities = 34/131 (26%), Positives = 55/131 (42%), Gaps = 23/131 (18%), Frame = +2

```
Query   1685   LQFSDCSMYPRVSPTPGREGTKTSPAFCSIDASTPIQRLALRSTAGPSGEEFSGVLHACI   1864
        + F + + Y VS      + T+  ++ S      Q+LA  + +G  E++
Sbjct   292   MYFQEANNYGLVSIIDGDKVSTRADYSYLS-----QQLASATPSGTTKSEYT-----   337

Query   1865   LRGQSCFSPQSPLIWCPPLRPWSATEDPLPPVASPTVLHLIFASLNCG-GDTEKYLQI-   2038
        P+ S L  CPP+  W AT  PLPP  +  +  + SL+C  DT  Q+
Sbjct   338   -----PTNSALQSCPPVNGDWLATASPLPPSPNGDLCSCMEDSLSCALKDTISDDQVE   389

Query   2039   LVLGDRCGIDI   2068
        + G  CG D+
Sbjct   391   KLFGTVCGYDV   401
```

Contig 11792

|XP_002999868.1| cellulose synthase catalytic subunit [*Verticillium albo-atrum* VaMs.102] Length=777

|XP_001259116.1| glycosyl transferase, putative [*Neosartorya fischeri* NRRL 181] Length=738

|XP_001273558.1| glycosyl transferase, putative [*Aspergillus clavatus* NRRL 1] Length=739

Score = 34.3 bits (77), Expect = 1E01 Identities = 16/28 (57%), Positives = 16/28 (57%), Gaps = 0/28 (0%), Frame = +3

```
Query   285   LGKFWVTALAPTPFAFWLVPEDFGGKTK   368
        LGK W TA      P F  VPEDFGG K
Sbjct   417   LGKGWKTAYIHEPLQFGTVPEDFGGHLK   444
```

Contig 12738

|YP_003637451.1| glycosyl transferase family 2 [*Cellulomonas flavigena* DSM 20109]

Length=335 Score = 39.3 bits (90), Expect = 7E-01, Identities = 25/45 (56%), Positives = 27/45 (60%), Gaps = 5/45 (11%), Frame = -1

```
Query   725   RSGAG---DWVAVAGVEPTVGDVEGFLGGESAELLREEGFRARGG   591
        R GAG  DW+AV G +  GDV GF GG  A LLR E  RA GG
Sbjct   147   RWGAGVDRDWLAVDGTQSPPGDVFGFHGG--AALLRTEAVRAVGG   189
```

Contig 24032

>ref|YP_003072318.1| glycosyltransferase, group 2 family protein [*Teredinibacter turnerae*]

Length=272, Score = 33.9 bits (76), Expect = 7E00 Identities = 13/15 (87%), Positives = 13/15 (87%), Gaps = 0/15 (0%), Frame = +1

```
Query   151   NGHNPCSWKADRLTV   195
        NGHN SCWKAD LTV
Sbjct   180   NGHNASCWKADALTV   194
```

Figure 11: Contig queries from the sequenced genome with partial identity to glycosyl transferases.

MKKNHGYEEVEIDVLQTPQHLLHHYHNNSVPQPYGGYGQPAPEYSSRPGTANGRQPQANPDEYYSVWEPAGTG
 RYSAQSFGESEPVHSRPETPSLNSAYGTPYDLRPPPKNVYSSDSGPTTPMQSQLSLAALLPNGPAPVIDADWV
 NKSNSIARLHDDRDDADIWKGWKRYVFKFVPVLTILNTAMYLlyLGyRIYCVVAAQKLrNTTYAQAWVFIGVEI
 AVALPALMHNIWTMMAMKKRLRPKLRLTSNDDV**PSVD**VFVTCCGEEDDLVADTVRAACNL**DYPR**DRFRVVVL**D**
DGKSEGLENAVLGMSQTYPNLV**YIAR**PKIPGKPH**HF****KAGNLN**YGLDAVHQLPGGAASSWPPWTLTWYIIPERDW
 LRAVLPHLLVDDKMALACPPQLFYNT**PDS**PLAQSLDFVHVIEPIKDALGVAWCTGSGYVVRREALDQIGNF
 PLGSLAE**D**VATSTLM**LGKGWKTAYI**HE**PLQ**FGTV**PEDFGGHL****KQTRW**AIGTVDTSFKLNFCWGDKVREMSF
 AQRFSGLYAMLSLYTVLLTISMFAIPIILVMGKPLVAFATDDQFRWLIRACFAATISNRLCEFALFIPAGYH
 TGQRGSRYQLWMAPYIALCIVRSFILPTWLGGQAQAFKPTGSLGSALNERDAHSRKNMMRRLWAILVNYMGLF
 HLGfVYlTLVGvLTsYRCfYLDtTVTDVLRCLVTHAFWPPLTFLFICSSlWTPVAYAIpPTMPEREALlDR
 DPkTGVAHPTRQRkKIAFGGQAawFELEYtFTTAYtALiFAASFFlF

Figure 12: CESA_CelA like cellulose synthase GT-2 sequence from *Verticillium albo-atrum*. The bold single underlined amino acids are partially conserved among β -1,3-glycosyl transferases. The double-underlined QXXRW is the glycosyl transferase processive motif.

*NSIEENAE*TGETQQRGPAAGRRLSGLPPTAQRRERAGSAVLERASQPAVAERRAAVQRARDGQPSPSRSED
 PRERLSVGGVQHAGVFKD*RD**LGKFW**TALAPT**PFAFWL****VPEDFGGKTK**VVPGFLGPSAIIHPHCKKTVLNVPK
 SLDA**QCGRW**PALRPKRGPGQGTVLNNKKLRSLLDPRPKMWNQETRVSGRCFLRCLKKA*K

Figure 13: Contig 11792. The highlighted sequences share identity with the CESA_CelA like cellulose synthase from *V. albo-atrum* where the bold amino acids show complete identity.

4.3.2.1 Contig 11792

Contig 11792 shows identity to the partially conserved fungal cellulose synthase sequence VPEDFGGXXK of *Verticillium albo-atrum*, *Aspergillus clavatus* and *Neosartorya fisheri* (Figure 11). The VPEDFGGXXK sequence is followed immediately by QTRWA (Figure 12), the highly conserved QXXRW β -glycosyl transferase processive motif. The *Verticillium albo-atrum* β -glycosyl transferase domain A consists of the the UDP-glucose binding DDG motif, the partially conserved KAG and highly conserved DXD motifs. Domain

B motifs include a single D amino acid motif and the QXXRW. Amino acids not highlighted but in bold are at least partially conserved among plant (Delmer, 1999), fungal (Ruiz-Herrera *et al.*, 2002; Beauvais *et al.*, 2001) and bacterial (Stasinopoulos *et al.*, 1999) β -1,3-glucan synthases. The amino acid sequence of contig 11792 has the QCGRW motif 28 amino acids downstream in frame with the VPEDFGGXXK sequence (Figure 13). The highlighted sections in bold are the amino acids which share identity between the contig 11792 and the *V. albo-atrum* subject sequences.

The nucleotide sequence was BLAST searched against NCBI *Euglena* EST's, but no EST showed significant identity. The presence of stop codons in this sequence and the lack of other confirmable motifs are most likely due to incorrect assembly of the contig. A cellulose synthase CelsA like gene product, producing a β -1,3/ β -1,4- glucan, was recently identified in *Aspergillus* with some identity to plant and bacterial β -1,3- and β -1,4- glycosyl transferases while being absent from other fungi such as *S. cerevisiae* (de Groot *et al.*, 2009). As neither *Euglena* (Pigon, 1947) or *Aspergillus* (de Groot *et al.*, 2009) produce cellulose and presence of the QXXRW motif and the VPEDFGGXXK sequence suggest that at least part of the sequence on contig 11792 may code for a β -1,3-glycosyl transferase involved in paramylon synthesis.

4.3.2.2 Contig 11487

Contig 11487 showed partial identity to *Aspergillus fumigatus* and *Neosartorya fisheri* β -1,3-glucanoyl transferases. The nucleotide sequence of the entire contig is 2494bp but only a small section (highlighted) shows partial identity to the β -1,3-glycosyl transferase sequences (Figure. 11; Figure. 14, highlighted). The amino acids in bold share identity with subject β -1,3-glucan synthase sequences in the +2 (Figure 14) reading frame. Interestingly, the β -1,3-glycosyl transferase KAG motif is partially conserved in the + 3 reading frame (Figure 15). The underlined section represents the location of the motif while the bold amino acids indicate shared identity. The KAG motif is only partially conserved amongst β -1,3-glycosyl transferases (Ruiz-Herrera *et al.*, 2002; Beauvais *et al.*, 2001; Stasinopoulos *et al.*, 1999), however the non-identical amino acids in the motif are substituted with amino acids of similar polarity and charge. It can be speculated that an incorrect assembly or changes in the amino acid sequence may have caused a frame shift in the contig. When contig 11487 was BLAST searched against the transcriptome for other contig which share nucleotide identity, two smaller contigs 10042 and 11670 were identified that showed similar nucleotide sequences (data not shown) but significant differences for the amino acid sequences based on substitutions and deletions indicating sequence misreads during assembly.

QWRLVALLSLSELPFFPMALPLFWQNHLSARPTDATALATVLSQA AVNPSTLVWPPVASALEECPPLGEQP*A
 AELRAAMGLAPNQGNFQYCGLPASLQSSGSGMDVFPNGLTVRIHLPPPLIFLWPSLFAVVSAFFATDAFADR
 PRG*AVYV*DALC*GSLCGMAEAPG*GI*WERPHWQSAVASITKCECSGGFLRDDPISHAPLHRLWRGGSCLA
 WAGSGPKRPRLSPQQPHREPYNGGPVYLLEALSPC*LSASPSTGRPAGLLGPLSPLGALSHLPRHGRRVRTPL
 PPREQSIRSPQPRPAVLRLWCLSLCPRRPPLASI*IPTIAVACTAGKSEFATNVQEPHHSSQLVLSHVENLFF
 SLKGPA CFQ*EANCSE*TFKEITKQNEVISGLHQTVGFCCLLVLQKCATYSMQL*RVD RDHSDTFV*RRHGAQ
 VFHVAF*SGLGRPR*ESHSQPSRLLSLDSQVRGLMANALSLISHRVRPKTRFMPPLFLVTLRPF SVVIRK GK
 H*HPCRTPRICRGYLWTEGGSAGSPRLPET*VM*SVFFSVFNAQNCLCL**QFSDCSMYPRVSP**TPGREGTKT
SPAFCSIDASTPI**QRLAL**RSTAGPSGEEFSGVLHACILRGQSCF**PSQSPLIWCPPLPRWSATEDPLPP**VASP
 TVLHLIFAS**SLNCGGDTEKYLQILVLGDRCGID**IMFLQEAGRPFFHFSFRVPSSSPAHHLEEEGGKQSWYEIPY
 *GCVCAKTRCIAIFHSDGFGWCFGGCLTLP TSSDIYLGKHLHSPGCFHGAWKQKRACISDLLWLRPAPLMAMG
 GFQRGSENSQATTPAVSCPWARGGMEFP

Figure 14: Contig 11487 ORF +3. Contig 11487 had partial identity to a putative β -1,3-glycosyltransferase of *N. fischeri* (Figure. 12). The highlighted section in bold are the amino acids sharing identity.

NGVWSHFFHYRSCPFQWPCRCSGKIIYRQGPPTQPWRLSCHRQL*TCHLRWSGLLWHPLWRNAHLWENNPER
 QSSGLRWV*PQINRATFSIAACRPRCSPLVLAWTFFRMASQ*GFIYHLPSFFYGLLFLP*CQHFLRPMHLLID
 QEARLFMSKMLYVRVPSAAWQKRLGEASDGSVPIGKALLPRSPNVSVVGVF*GMTPSAMLPSTASGVVGAA*H
 GQALDPSVLD*ALSSPIASRTMGDRFTSSRHCPPADFPARALGDQRDFSARCLPSALFLTYPGTGAGFELHC
 LLASSRFVPPSHGLRYCVGSAPFCAPADRPWRA**YESQ**QS**SHARP****GNLN**LPRTFKNPTILPNLFFLTWRICSF
 LSRVPHAFSERLTVLNEHLRKLPSKMKS FQVSIKPLGSAACWFCKNVQHIPCNCSEWTAIIQTPLCNADMGPR
 YFMWLFEVVWAGHVESHIRNPPAACSLWTAKSVV*WQMLFHSSVTGCVPKPVSCRPCFWLPSGHFLL*SGKGS
 IDIRAAPHAFVDTFGRREVQLDHPRDCRPKPE*CNLFFFRYSMHKIVYVYNFLTVLCIRGSAPRLGEKAPKR
 RLHFVASMPQLLSND*PSDLQPDLLGRNFQGSSTHAF*EVNHASRPSHL*YGV LHCHAHGQPRKTHSLLLHPQ
 LFSI*FLPV*IVAVTQKNICRF*CLEIDAV*TLCSSKRPDVHFFFTASEYQVLLLHTIWKRGGSNPGTKFHI
 EVVCVQKPGASQFSTVMGLDGVLVVVLHFPQAVTFILANIYIPQGVFMALGSKKGHAFLIFSG*GQHL*WQWG
 DFNEEARTAKPPRRLYRARGPEGAWSS

Figure 15: Contig 11487 ORF +2. Contig 11487 in ORF +2 showed amino acids with identity (bold underlined) to the partially conserved β -1,3-glucan domain A KAG motif a frame shift may have been caused by incorrect assembly.

General discussion, conclusion and future work

To date, research on *Euglena* has focused mainly on phylogeny and the physical and morphological characteristics of the organism and its novel β -1,3-glucan storage polymer, paramylon. The most recent publication (Bäumer *et al.*, 2001) identified two peptides, within the 670 kDa membrane-bound paramylon synthase complex, approximately 54 and 37 kDa with affinity for the substrate, UDP-glucose, as well as increased activity with Ca^{2+} as the co-factor similar to what has been observed in plant callose synthases. These results have been confirmed in our laboratory (van der Merwe, 2007). However, none of the genes involved in paramylon synthesis have been identified. In this study, several approaches were followed in an attempt to further identify genes involved in paramylon synthesis in *E. gracilis*.

Screening of a *E. gracilis* cDNA library using a polyclonal antibody and radio-labeled substrate did not identify any sequences encoding subunits of paramylon synthase, possibly due to the membrane-bound β -1,3-glycosyl transferases losing conformational stability when expressed in *E. coli*. Protein purification steps identified peptides associated to the paramylon membrane and involved in paramylon synthesis. In this study nine proteins were identified using a polyclonal antibody and the corresponding peptides were excised, digested and the sequences predicted by LCMSMS. The protein bands corresponded with those previously published (Bäumer *et al.*, 2001). Five other peptides were also identified by the polyclonal antibody, possibly transport or structural proteins or β -1,3-glycosyl hydrolases associated to the membrane. Predicted peptides were used to probe a partially sequenced transcriptome sequenced in this study as well as an uncompleted genome from another study. As glycosyl transferase sequences are highly variable, only predicted peptides matching a highly conserved sequence or motif would be a useful probe. However, none of the predicted peptides identified putative glycosyl transferases from the genome or transcriptome.

The sequenced genome was screened using BLASTx for putative glycosyl transferase enzymes, specifically for glycosyl transferase GT-2 and β -1,3-glucan synthases. Candidate glycosyl transferases sharing specific domain structures and partial sequence identity to β -1,3-glucan synthase, cellulose synthase genes and glycosyl transferase GT-2 sequences were identified. Although all the listed sequences are potential candidates, two candidate sequences drew specific attention. A sequence on contig 11487 showed identity to *Neosartorya fisheri* and *Aspergillus fumigatus* β -1,3-glycosyl transferase. A KAG-like β -1,3-glycosyl transferase conserved motif lies upstream of the region showing identity to the β -1,3-glycosyl transferase (Stasinopoulos *et al.*, 1999) but out of frame, likely due to missassembly of the contig. A sequence on contig 11792 showed similarity to a conserved cellulose synthase-like sequence of *Verticillium albo-atrum*, *Aspergillus fumigatus* and *Neosartorya fisheri*. Both the contig sequence and the subject sequences had a QXXRW motif (Delmer, 1999; Ruiz-Herrera *et al.*, 2002; Beauvais *et al.*, 2001), the processive motif, down-

stream and in-frame with the sequence showing identity. Interestingly, the query sequences showed identity to subject sequences of the same fungal genera. As these fungal genera do not synthesize cellulose but make a β -1,3-/ β -4- glucan from a celA cellulose synthase-like gene product of a similar molecular weight (de Groot *et al.*, 2009), it can be speculated that the paramylon synthase genes may have similar sequence and structural characteristics.

Future work would include amplifying and cloning all candidate sequences from cDNA using primers based on the sequence information of the contigs. Cloned sequences would need to be expressed and screened for activity.

In conclusion, subunits of the paramylon synthase complex of *E. gracilis* were partially purified, unidentified proteins associated to the granular membrane were described and candidate β -1,3-glycosyl transferase nucleotide sequences were identified.

Reference list

- Adl, S.M. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol.* **52**: 399–451
- Ahmadinejad, N., Dagan, T. and Martin, W., (2007) Genome history in the symbiotic hybrid *Euglena gracilis*. *Gene* **402**: 35-39
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**: 403 - 410
- Archibald, J.M., Rogers, M.B., Toop, M., Ishida, K. and Keeling, P.J. (2003) Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. *Proc. Natl. Acad. Sci. U. S. A.* **100**: 7678–7683
- Arellano, M., Duran, A. and Pérez, P. (1996). Rho1 GTPase activates the (1,3)- β -glucan synthase and is involved in *Schizosaccharomyces pombe* morphogenesis. *EMBO J.* **15**: 4584-4591
- Atkinson, E.M. and Long, S. (1992) Homology of *Rhizobium meliloti* NodC to polysaccharide polymerizing enzymes. *Mol. Plant-Microbe Interact.* **5**: 439–442
- Auer, M., Scarborough, G.A. and Kuhlbrandt, W. (1998) Three-dimensional map of the plasma membrane HC-ATPase in the open conformation. *Nature* **392**: 840–42
- Aswald, P.D., Frost, D., Drake, R.R. and Selitrennikoff, C.P. (1994) (1,3)- β -Glucan synthase activity of *Neurospora crassa*: Identification of a substrate-binding protein, *Biochim. Biophys. Acta* **1201**: 312-320
- Aswald, P., Zugel, M., Mones, C., Frost, D. and Selitrennikoff, C.P. (1993) Purification of 1,3- β -glucan synthase from *Neurospora crassa* by product entrapment. *Exp. Mycol.* **17**: 130-141
- Baba M., Nakajima, M., Schols, D., Pauwels, R., Balzarini, J. and De Clercq, E. (1988) Pentosan polysulfate, a sulfated oligosaccharide, is a potent and selective anti-HIV agent in vitro. *Antiviral Res.* **9**: 335-343
- Baba, M., Schols, D., De Clercq, E., Pauwels, R., Nagy, M., Györgyi-Edelényi, J., Löw, M. and Görög, S. (1990) Novel sulfated polymers as highly potent and selective inhibitors of human immunodeficiency virus replication and giant cell formation. *Antimicrob. Agents Chemother.* **34**: 134-138

- Barras, D.R., Stone, B.A. (1968) Carbohydrate composition and metabolism in *Euglena*. In Buetow DE (ed.), *The Biology of Euglena*, vol. II, Biochemistry. Academic Press, New York, pp. 149–191
- Baskin, T.I., Betzner, A.A., Hoggart, R., Cork, A. and Williamson, R.E. (1992) Root morphology mutants in *Arabidopsis thaliana*. *Aust. J. Plant Physiol.* **19**: 427–37
- Bäumer, D., Preisfeld, A. And Ruppel, H.G. (2001) Isolation and characterization of paramylon synthase from *Euglena gracilis* (Euglenophyceae). *Journal of Phycology* **37**: 38-46
- Beauvais, A.R., Bruneau, J.M., Mol, P.C., Buitrago, M.J., Legrand, R. and Latgé, J.P. (2001) Glucan synthase complex of *Aspergillus fumigatus*, *Journal of Bacteriology* **183**: 2273-2279
- Beauvais, A., Drake, R., Ng, K., Diaquin, M. and Latge, J. P. (1993) Characterization of the 1,3- β -glucan synthase of *Aspergillus fumigatus*, *J. Gen. Microbiology* **139**: 3071 -3078
- Besemer, J. and Borodovsky, M. (1999) Heuristic approach to deriving models for gene finding. *Nucleic acids Research* **27**: 3911-3920
- Blaurock, A.E. and Stoeckenius, W. (1971) Structure of the purple membrane. *Nature* **233**: 152-155
- Booy, F.P., Chanzy, H. and Boudet, A. (1981) An electron diffraction from the primary wall of cotton fibers. *Protoplasma* **94**: 299–306
- Bourne, Y. and Henrissat, B. (2001) Glycoside hydrolases and glycosyltransferases: families and functional modules. *Curr. Opin. Struct. Biol.* **11**: 593–600
- Briand, J. and Calvayrac, R. (1980) Paramylon synthesis in heterotrophic and photoheterotrophic *Euglena* (Euglenophyceae). *J. Phycol.* **16**: 234–239
- Brown, G.D. and Gordon, S. (2003) Fungal β -glucans and Mammalian immunity. *Immunity* **19**: 311-315
- Bulawa, C. E. (1992) *CSD2*, *CSD3*, and *CSD4*, genes required for chitin synthesis in *Saccharomyces cerevisiae*: the *CSD2* gene product is related to chitin synthases and to developmentally regulated proteins in *Rhizobium* species and *Xenopus laevis*. *Mol. Cell. Biol.* **12**: 1764–1776
- Bulone, V., Girard, V. and Fevre, M. (1990) Separation and partial purification of β -1,3- glucan and β -1,4- glucan synthase from *Saprolegnia*, *Plant Physiology*, **94**: 1748-1755
- Cabib, E. and Kang, M. S. (1987) Fungal 1,3-p-glucan synthase. *Methods Enzymol.* **138**: 637-642
- Campbell, J.A., Davies, G.J., Bulone, V. and Henrissat, B. (1997) A classification of nucleotide-diphosphosugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.* **326**: 929–939

- Carpita, N.C. (1996) Structure and biogenesis of the cell walls of grasses. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**:445–476
- Cavalier-Smith, T. (2003a) Protist phylogeny and the high-level classification of Protozoa. *Eur. J. Protistol.* **39**: 338–348
- Cavalier-Smith, T. (2003b) The excavate protozoan phyla Metamonada Grasse' emend. (Anaeromonadea, Parabasalia, *Carpodimonas*, Eopharyngia) and Loukozoa emend. (Jakobea, *Malawimonas*): their evolutionary affinities and new higher taxa. *Int. J. Syst. Evol. Microbiol.* **53**: 1741–1758
- Coutinho, P.M., Deleury, E., Davies, G.J., Henrissat, B. (2003) An Evolving Hierarchical Family Classification for Glycosyltransferases. *J. Mol. Biol.* **328**: 307–317
- Currier, H.P. (1957) Callose substance in plant cells. *Am J Bot* **44**: 478-488
- Dagan, T. and Martin, W. (2006) The tree of one percent. *Genome. Biol.* **7**: 118
- Daniel, M.D., Letvin, N.L., King, N.W., Kannagk M., Sehgal, P.K. and Hunt. R.D. (1985) Isolation of T-cell tropic HTLV III-like retrovirus from Macaques. *Science* **228**: 1201-1204
- Dawson, N.S. and Walne, P.L. (1991) Structural characterization of *Eutreptia pertyi* (Euglenophyta) I. General description. *Phycologia* **30**: 28-302
- De Angelis, P. L., Yang N. and Weigel, P. H. (1994) The *Streptococcus pyogenes* hyaluronan synthase: sequence comparison and conservation among various group A strains. *Biochem. Biophys. Res. Commun.* **199**: 1–10
- De Clercq, E. (1986) Chemotherapeutic approaches to the treatment of the acquired immune deficiency syndrome (AIDS). *J. Med. Chem.* **29**: 1561-1569
- De Groot, P.W.J., Brandt, B.W., Horiuchi, H., Ram, A.F.J., de Koster, C.G., Klis, F.M. (2009) A comprehensive genomic analysis of cell wall genes in *Aspergillus nidulans*, *Fungal Genetics and Biology*, **46**: S72-S81
- De Madariaga, I. (1992) Interspecific differences in the photosynthetic carbon metabolism of marine phytoplankton. *Mar. Biol.* **114**: 509–515
- Debellé, F., Rosenberg, C. and Denarie', J. (1992) The *Rhizobium*, *Bradyrhizobium*, and *Azorhizobium* NodC proteins are homologous to yeast chitin synthases. *Mol. Plant-Microbe Interact.* **5**:443–446

- Delmer D.P., Heiniger, V. and Kulow, C. (1977) UDP-glucose: glucan synthetase in developing cotton fibres. *Plant physiol* **59**: 713-718
- Delmer, D.P. (1983) Biosynthesis of cellulose. *Adv Carbohydr Chem Biochem* **41**: 105-153
- Delmer, D.P. (1999) Cellulose Biosynthesis, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**: 245–76
- Delmer, D.P., Cooper, G., Alexander, D., Cooper, J., Hayashi, T., Nitsche, L. and Thelen, M. (1985) New approaches to the study of cellulose synthesis. *J Cell Sci Suppl* **2**: 33-50
- Delmer, D.P., Solomon, M. and Read, S.M. (1991) Direct photolabeling with [³²P]UDP-glucose for identification of a subunit of cotton fiber callose synthase. *Plant Physiol.* **95**: 556–563
- Deslandes, Y., Marchessault, R. H., and Sarko, A. (1980) Triple-helical structure of (1,3)- β -D-glucan. *Macromolecules* **13**: 1466–1471
- Dhugga K.S. and Ray P.M. (1998) A 55 kDa plasma membrane-associated polypeptide is involved in β -1,3-glucan synthase activity in pea tissue. *FEBS Lett.* **278**(2): 283–286
- Dhugga, K.S. and Ray, P.M. (1994) Purification of 1,3- β -D-glucan synthase activity from pea tissue. *Eux J. Biochem.* **220**, 943-953
- Diaz, M., Sanchéz, Y., Bennet, T., Sun, C.R., Godoy, C., Tamanoi, F., Duran, A. and Pérez, P. (1993) The *Schizosaccharomyces pombe* *cwg2+* gene codes for the β -subunit of a geranylgeranyltransferase type I required for β -glucan synthesis. *EMBO J.* **12**: 5245-5254
- Dougherty, B.A. and van de Rijn, I. (1994) Molecular characterization of *hasA* from an operon required for hyaluronic acid synthesis in Group A *Streptococci*. *J. Biol. Chem.* **269**: 169–175
- Droege ,M. and Hill, B. (2008) The Genome Sequencer FLX™ System—longer reads, more applications, straight forward bioinformatics and more complete data sets. *Journal of Biotechnology.* **136**: 3 – 10
- Enderlin, C.S. and Selitrennikoff, C.P. (1994) Cloning and characterization of a *Neurospora crassa* gene required for (1,3)- β -glucan synthase activity and cell wall formation. *Proc. Nut1 Acad. Sci. USA* **91**: 9500-9504
- Eschrich, W. (1975) Sealing systems in phloem. *Encyclopaedia of Plant Physiology, transport in Plants*, I. Springer-Verlag, Berlin, pp 39-56

- Fenchel, T., Bernhard, C., Esteban, G., Finlay, B.J., Hansen, P.J. and Iversen, N. (1995) Microbial diversity and activity in a Danish fjord with anoxic deep water. *Ophelia* **43**: 45–100
- Fincher, G.B. and Stone, B.A. (1981) Metabolism of noncellulosic polysaccharides. Encyclopaedia of Plant Physiology, Plant Carbohydrates, II. Springer-Verlag, Berlin, pp 68-132
- Fink, J., Jeblick, W. and Kaus, H. (1990) Partial purification and immunological characterization of 1,3- β -glucan synthase from suspension cells of *Glycine max*. *Planta* **181**: 343-348
- Frost, D.J., Read, S.M., Drake, R.R., Haley, B.E. and Wasserman, B.P. (1990) Identification of the UDP-glucose-binding polypeptide of callose synthase from *Beta vulgaris* L. by photoaffinity labelling with 5-azido-UDP-glucose. *J. Biol. Chem.* **265**: 2162–2167
- Fukuchi, K., Sakagami, H., Ikeda, M., Kawazoe, Y., Oh-hara, T., Konno, K., Ichikawa, S., Hata, N., Kondo, H. and Nonoyama, M. (1989) Inhibition of herpes simplex virus infection by pine cone antitumor substances. *Anticancer Res.* **9**: 313-318
- Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J. P. (1987) Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. *FEBS Lett.* **224**: 149–155
- Gardner, K.H. and Blackwell, J. (1971) The substructure of cellulose microfibrils from the cell wall of the algae, *Valonia ventricosa*. *J. Ultrastruct. Res* **36**: 725-731
- Gibbs, S.P. (1978) The chloroplast of *Euglena* may have evolved from symbiotic green algae. *Can. J. Bot.* **56**: 2883–2889
- Girard, V. and MacLachlan, G. (1987) Modulation of pea membrane β -glucan synthase activity by calcium, colocation, endogenous protease and protease inhibitor. *Plant Physiol.* **85**: 131-136
- Glucksmann, M. A., Reuber, T. L. and Walker, G. C. (1993) Family of glycosyl transferases needed for the synthesis of succinoglycan by *Rhizobium meliloti*. *J. Bacteriol.* **175**: 7033–7044
- Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**: 95-98
- Harada, H., Sakagami, H., Nagata, K., Oh-hara, T., Kawazoe, Y., Ishihama, A., Hata, N., Misawa, Y., Terada, H. and Konno, K. (1991) Possible involvement of lignin structure in anti-influenza virus activity. *Antiviral Res.* **15**: 41-50
- Hayashi, M., Toda, K., Yoneji, T., Sato, O. and Shozaburo, K. (1993) Dietary value of rotifers and *Artemia* enriched with *Euglena gracilis* for red sea bream. *Nippon Suisan Gakkaishi* **59**: 1051–1058

- Hayashi, T., Read, S.M., Bussell, J., Thelen, M., Lin, F.C., Brown, R.M. Jr. and Delmer, D.P. (1987) UDP-glucose: (1-3)- β -glucan synthases from mung bean and cotton. *Plant Physiol.* **83**: 1054–1062
- Henderson, R. and Unwin, P.N.T. (1975) Three-dimensional model of purple membrane obtained by electron microscopy. *Nature (Lond.)* **257**: 28-32
- Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**: 309–316
- Henrissat, B. and Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**: 781–788
- Henrissat, B. and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **316**: 695–696
- Henze, K., Badr, A., Wettern, M., Cerff, R. and Martin, W. (1995) A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. *Proc. Natl. Acad. Sci. U. S. A.* **92**: 9122–9126
- Hietä, K., Kuga, S., Usuda, M. (1984) Electronstaining of reducing ends evidences a parallel-chain structure in Valonia Cellulose *Biopolymers* **23**: 1807-1810
- Hill, R.L. and Brew K. 1975. Lactose synthetase. *Adv. Enzymol. Rel. Areas Mol. Biol.* **43**: 411–517
- Hirano, T., Taga, T., Nakano, N., Yasukawa, K., Kashiwamura, S., Shimizu, K., Nakajima, K., Pyun, K.H. and Kishimoto, T. (1985) Purification of homogeneity and characterization of human B-cell differentiation factor (BCDF or BSF-2). *Proc. Natl. Acad. Sci. U.S.A.*, **82**: 5490-5494
- Hirosawa, M., Hoshida, M., Ishikawa, M. and Toya, T. (1993) MASCOT: multiple alignment system for protein sequences based on three-way dynamic programming. *Computer Applications in the Biosciences*, **9**: 161-167
- Ho, S.N., Abraham, R.T., Nilson, A., Handwerger, B.S. and McKean, D.J. (1987) Interleukin 1-mediated activation of interleukin 4 (IL-4)-producing T lymphocytes. Proliferation by IL-4-dependent and IL-4 independent mechanisms. *J. Immunol.*, **139**: 1532-1540
- Hong, Z., Mann, P., Brown, N.H., Tran, L.E., Shaw, K.J., Hare, R.S. and DiDomenico, B. (1994) Cloning and characterization of *KNR4*, a yeast gene involved in (1,3)- β -glucan synthesis. *Mol. Cell. Biol.* **14**: 1017-1025

Hu, Y. and Walker, S. (2002) Remarkable structural similarities between diverse glycosyltransferases. *Chem. Biol.* **9**: 1287–1296

Hundle, B.S., O'Brien, D.A., Alberti, M., Beyer, P. and Hearst, J.E. (1992) Functional expression of zeaxanthin glucosyltransferase from *Erwinia herbicola* and a proposed uridine diphosphate binding site. *Proc. Natl. Acad. Sci. USA* **89**: 9321–9325

Ikebuchi, K., Wong, G.G., Clark, S.C., Ihle, J.N., Hirai, Y. and Ogwa, M. (1987) Interleukin 6 enhancement of interleukin 3-dependent proliferation of multipotential homopoietic progenitors. *Proc. Natl. Acad. Sci. U.S.A.*, **84**: 9035-9039

Inoue, S.B., Takewake, N., Takasuka, T., Mio, T., Adachi, M., Fujii, Y., Miyamoto, C., Arisawa, M., Furuichi, Y. And Watanabe, T. (1995) Characterization and gene cloning of β -1,3-glucan synthase from *Saccharomyces cerevisiae*. *European Journal of Biochemistry* **231**: 845-854

Inoue, S.B., Qadota, H., Arisawa, M., Anraku, Y., Watanabe, T. and Ohya, Y. (1996) Signalling toward yeast β -1,3-glucan synthesis. *Cell Struct. Funct.* **21**: 395-402

Inoue, S.B., Qadota, H., Arisawa, Y., Watanabe, T. and Ohya, Y. (1999) Prenylation of Rho1 is required for activation of yeast β -1,3-glucan synthase. *J. Biol. Chem.* **274**: 38113-38124

Inui, H., Miyatake, I., Nakanoy, Y. and Kitaokas, K. (1983) Production and composition of wax esters by fermentation of *Euglena gracilis*. *Agricultural and Biological Chemistry* **47**: 2669-2671

Ito, M., Baba, M., Sato, A., Pauwels, R., De Clercq, E. and Shigeta, S. (1987) Inhibitory effect of dextran sulfate and heparin on the replication of human immunodeficiency virus (HIV) in vitro. *Antiviral Res.* **7**: 361-367

Itoh, W., Sugawara, I., Kimura, S., Tabata, K., Hirata, A., Kojima, T., Mori, S. and Shimada, K. (1990) Immunopharmacological study of sulfated Schizophyllan (SPG). I. Its action as a mitogen and anti-HIV agent. *Int. J. Immunopharm.* **12**: 225 233

Jacobson, R.H., Zhang, X.-J., DuBose, R.F. and Matthews B.W. (1994) Three-dimensional structure of β -galactosidase from *E. coli*. *Nature (London)* **369**: 761–766

Kasahara, S., Yamada, H., Mio, T., Shiratori, Y., Miyamoto, C., Yabe, T., Nakajima, T., Ichishima, E. and Furuichi, Y. (1994a) Cloning of the *Saccharomyces cerevisiae* gene whose overexpression overcomes the effects of HM-1 killer toxin, which inhibits β -glucan synthesis. *J. Bacteriol.* **176**: 1488-1499

- Kato, K. (1981) Ultrastructure of the plant cell wall biochemical viewpoint. In W. Tanner and F.A.Loewus [eds.], Plant carbohydrates II. Extracellularcarbohydrates, pp 29-46. Springer-Verlag, Berlin, Heidelberg, New York.
- Kauss, H. and Jeblick, W. (1985) Callose biosynthesis as a Ca^{2+} regulated process and possible relations to the induction of other metabolic changes. *J Cell Sci* **2**: Suppl, 98-103
- Kauss, H. and Jeblick, W. (1986) Synergistic activation of β -1,3-glucan synthase by Ca^{2+} and polyamines. *Plant Sci* **43**: 103-107
- Kauss, H., Kohle, H. and Jeblick, W. (1983) Proteolytic activation and stimulation Ca^{2+} of glucan synthase from soybean cells. *FEBS Lett* **158**: 84-88
- Kelly, R., Register, E., Hsu, M-J., Kurtz, M. and Nielsen, J. (1996) Isolation of a gene involved in 1,3- β -glucan synthase in *Aspergillus nidulans* and purification of the corresponding protein. *J. Bacteriol.* **178**: 4381–4391
- Kiss, J.Z. and Triemer, R.E. (1988) A comparative study of the storage carbohydrate granules from *Euglena* (Euglenida) and *Pavlova* (Prymnesiida). *J. Protozool.* **35**: 237–41
- Kiss, J.Z., Vasconcelos, A.C. and Triemer, R.E. (1986) Paramylon synthesis and chloroplast structure associated with nutrient levels in *Euglena* (Euglenophyceae). *J. Phycol.* **22**: 327–333
- Kiss, J.Z., Vasconcelos, A.C. and Triemer, R.E. (1988) The intramembranous particle profile of the paramylon membrane during paramylon synthesis in *Euglena* (Euglenophyceae). *J. Phycol.* **24**: 152–157
- Kiss, J.Z., Vasconcelos, A.C., Triemer, R.E. (1987) Structure of the euglenoid storage carbohydrate, Paramylon. *Amer. J. Bot.* **74**(6): 877-882
- Kitaoka, S. and Hosotani, K. (1977) Studies on culture conditions for the determination of the nutritive value of *Euglena gracilis* protein and the general amino acid composition of the cells. *Nogei-kagaku* **51**: 477–482
- Koizumi, N., Sakagami, H., Utsumi, A., Fujinaga, S., Takeda, Minoru, Asano, K., Sugawara, I., Ichikawa, S., Kondo, H., Mori, S., Miyatake, K., Nakano, Y., Nakashima, H., Murakami, T., Miyano, N. and Yamamoto, N. (1993) Anti-HIV (human immunodeficiency virus) activity of sulfated paramylon. *Antiviral Research*, **21**: 1-14
- Kojima, T., Tabata, K., Hirata, A. and Sugawara, I. (1986) Biological activity of sulfated schizophyllan. *Agric. Biol. Chem.* **50**: 1635–1636
- Kondo, Y., Kato, A., Hojo, H., Nozoe, S., Takeuchi, M. and Ochi, K. (1992) Cytokine-related immunopotentiating activities of paramyon, a β -(1,3)-D-glucan from *Euglena gracilis*. *J. Pharmacobio-Dyn.* **15**: 617-621

- Kondoh, O., Tachibana, Y., Ohya, Y., Arisawa, M. and Watanabe, T. (1997) Cloning of *Rhol* gene from *Candida albicans* and its regulation of β -1,3-glucan synthesis. *J. Bacteriol.* **179**: 7734-7741
- Kreeger, D. R. and Van der Veer, J. (1970) Paramylon in a chrysophyte. *Act. Bot. Neerl.* **19**: 401-402
- Kreimer, G. (1994) Cell biology of phototaxis in flagellate algae. *Int. Rev. Cytol.* **148**: 229-310
- Kudlicka K, Brown RM Jr. (1997) Cellulose and callose biosynthesis in higher plants. I. Solubilization and separation of (β -1,3)- and (β -1,4)-glucan synthase activities from mung bean. *Plant Physiol.* **115**: 643-56
- Kusmic, C., Barsacchi, R., Barsanti, L., Gualtieri, P. and Passarelli, V. (1999) *Euglena gracilis* as source of the antioxidant vitamin E. Effects of culture conditions in the wild strain and in the natural mutant WZSL. *J appl. Phycol.* **10**: 555-559
- Laine, R.A. (1994) A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: the isomer barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology*, **4**: 759-767
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 68-685
- Lawson, S.G., Mason, T.L, Sabin, R.D., Sloan, M.E., Drake, R.R., Boyd, H.E. and Wasserman, B.P. (1989) UDP-Glucose: (1,3)- β -Glucan Synthase from *Daucus carota* L, *Plant Physiol.* **90**: 101-108
- Leedale, G.F. (1967) *Euglenoid Flagellates*, Prentice-Hall, Englewood Cliffs, NJ, p. 242.
- Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. and Mornon, J.-P. (1990) Hydrophobic cluster analysis: procedures to derive structural and functional information from 2D-representation of protein sequences. *Biochimie* **72**: 555-574
- Li, L. and Brown, R.M. Jr. (1993) β -Glucan synthesis in the cotton fiber. II. Regulation and kinetic properties of β -glucan synthases. *Plant Physiol.* **101**: 1143-1148
- Li, L., Drake, R.R., Clement, S. and Brown, R.M. Jr. (1993) β -Glucan synthesis in the cotton fiber III. *Plant Physiol.* **101**: 1149-1156
- Lowenthal, J.W., Cerottini, J.C. and MacDonald, H.R., (1986) Interleukin 1-dependent induction of both interleukin 2-secretion and interleukin 2-receptor expression by thymoma cells, *J. Immunol.*, **137**: 1226-1231

- Manabe, H., Sakagami, H., Ishizone, H., Kusano, H., Fujimaki, M., Wada, C., Komatsu, N., Nakashima, H., Murakami, T. and Yamamoto, N. (1992) Effects of Catuaba extracts on microbial and HIV infection. *In Vivo* **6**: 161-166
- Marchessault, R. H. and Deslandes, Y. (1979) Fine structure of (1,3)- β -D-glucans: Curdlan and Paramylon. *Carbohydrate Res.* **75**: 231–242
- Marechal, L.R. and Goldemberg, S.H. (1964) Uridine diphosphate glucose: β -1,3-glucan β -glucosyltransferase from *Euglena gracilis*. *J. Biol. Chem.* **239**: 3163–3167
- Martin, W., Brinkmann, H., Savonna, C. and Cerff, R. (1993) Evidence for a chimeric nature of nuclear genomes — eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 8692–8696
- Mazur, P. and Baginsky, W. (1996) In vitro activity of 1,3- β -D-glucan synthase requires the GTP-binding protein Rho1, *J Biol Chem* **271**: 14604-14609
- McCormack, B.A., Gregory, A.C.E., Kerry, M.E., Smith, C. and Bolwell, G.P. (1997) Purification of an elicitor-induced glucan synthase (callose synthase) from suspension cultures of French bean (*Phaseolus vulgaris* L.): purification and immunolocalization of a probable Mr–65 000 subunit of the enzyme. *Planta* **203**: 196–203
- McFadden, G.I. (2001) Primary and secondary endosymbiosis and the origin of plastids. *J. Phycol.* **37**: 951–959
- Meikle, P.J., Ng, K.F., Johnson, E., Hoogenraad, N. J. and Stone, B.A. (1991) The p-glucan synthase from *Lolium multzporum*. *J. Biol. Chem.* **266**: 22569-22581
- Mitsuya, H., Looney, D.J., Kuno, S., Ueno, R., Wong-Staal, F. and Broder, S. (1988) Dextran sulfate suppression of viruses in the HIV family: inhibition of virion binding to CD4 + cells. *Science* **240**: 646-649
- Mizumoto, K., Sugawara, 1., Itoh, W., Kodama, T., Hayami, M. and Mori, S. (1988) Sulfated homopolysaccharides with immunomodulating activities are more potent anti-HTLV-III agents than sulfated heteropolysaccharides. *Jpn. J. Exp. Med.* **58**: 145-151
- Mueller, S. and Brown, R.M. (1980). The control of cellulose microfibril deposition in the cell wall of higher plants. I. Can directed membrane flow orient cellulose microfibrils? Indirect evidence from freeze fractured plasma membranes of maize and pine seedlings. *Platzta (Brrl.)* **154**: 489-500
- Mueller, S. and Machlachlan, G.A. (1983) Radioautographic visualisation of β -glucans formed by pea membranes from UDP-glucose. *Can J Bot* **61**: 1266-1275

- Muraguchi, A., Hirano, T., Tang, B., Matsuda, T., Horii, Y., Nakajima, K., Koyama, K. and Kishimoto, T. (1988) The essential role of B-cell stimulatory factor (BSF-2/IL-6) for the terminal differentiation of B cells, *J.Exp. Med.*, **167**: 332-344
- Nakashima, H., Murakami, T., Yamamoto, N., Naoe, T., Kawazoe, Y., Konno, K. and Sakagami, H. (1992a) Lignified materials as medicinal resources. V. Anti-HIV (human immunodeficiency virus) activity of synthetic lignins. *Chem. Pharm. Bull.* **40**: 2102-2105
- Nakamura, Y., Gojobori, T. and Ikemura, T. (1997) Codon usage tabulated from the international DNA sequence databases, *Nucleic Acids Research*, **26**: 334
- Nakashima, H., Murakami, T., Yamamoto, N., Sakagami, H., Tanuma, S., Hatano, T., Yoshida, T. and Okuda, T. (1992b) Inhibition of human immunodeficiency viral replication by tannins and related compounds. *Antiviral Res.* **18**: 91-103
- Nakashima, H., Yoshida, O., Baba, M., De Clercq, E. and Yamamoto, N. (1989) Anti-HIV activity of dextran sulphate as determined under different experimental conditions. *Antiviral Res.* **11**: 233-246
- Nakashima, H., Yoshida, O., Tochikura, T.S., Yoshida, T., Mimura, T., Kido, Y., Motoki, Y., Kaneko, Y., Uryu, T. and Yamamoto, N. (1987) Sulfation of polysaccharides generates potent and selective inhibitors of human immunodeficiency virus infection and replication *in vitro*. *Jpn. J. Cancer Res. (Gann)* **78**: 1164-1168
- Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-Coffee: A novel Method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology.* **302**: 205 - 217
- O'Brien, E.A., Koski, L.B., Zhang, Y., Yang, L., Wang, E., Gray, M.W., Burger, G. and Lang, B.F. (2007) TBestDB: a taxonomically broad database of expressed sequence tags (ESTs) *Nucleic Acids Research* (Database issue) **35**: 445 - 451
- Patterson, D.J. (1999) The diversity of eukaryotes. *Am Nat* **154** (Oct. Suppl.): 96–124
- Paulson, J.C. and Colley, K.J. (1989) Glycosyltransferases: Structure, localization and control of cell type-specific glycosylation. *J. Biol. Chem.* **264**: 17615–17618
- Pedersen, L.H., Jacobsen, S., Hejgaard, J. and Rasmussen, S.K. (1993) Characterization and partial purification of β -1,3-glucan (callose) synthase from barley (*Hordeum vulgare*) leaves, *Plant Science* **91**: 127-138
- Pigon, A., (1947) On the pellicle of *Euglena viridis* Ehrbg. *Bull. int. Acad. pol. Sci. B. H.*, 111-120.

Preston, R.D. (1964) Structural and mechanical aspects of plant cell walls with particular reference to synthesis and growth. In *The Formation of Wood in Forest Trees*, ed. MH Zimmermann, pp. 169–201. New York: Academic

Quader, H. (1986) Cellulose microfibril orientation in *Oocytis solitaria*, proof that microtubules control the alignment of the terminal complexes. *J. Cell Sci.* **83**: 223-234

Quadota, H., Python, C.P., Inoue, S.B., Arisawa, M., Anraku, Y., Zheng, Y., Watanabe, T., Levin, D.E. and Ohya, Y. (1996) Identification of yeast Rho1p GTPase as regulatory subunit of β -1,3-glucan synthase. *Science* **272**: 279-281

Quentmeier, H., Ingold, E. and Seitz, H.U. (1987) Purification of an autocatalytic protein-glycosylating enzyme from cell suspensions of *Daucus carota* L. *Planta* **171**: 483-488

Quesada, L.A., de Lusting, E.S., Marechal, L.R and Belocopitow, E. (1976) Antitumor activity of paramylon on sarcoma-180 in mice. *Gann* **67**: 455-459

Reyes-Prieto, A., Hackett, J.D., Soares, M.B., Bonaldo, M.F. and Bhattacharya, D. (2006) Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. *Curr. Biol.* **16**: 2320–2325

Robbins, E.A. and Steele, R.D. (1977) Cholesterol lowering effect of dietary yeast and yeast fractions. *J Food Sci.* **42**: 694–698

Roemer, T. and Bussey, H. (1991) Yeast p-glucan synthesis: *KRE6* encodes a predicted type II membrane protein required for glucan synthesis *in vivo* and for glucan synthase activity *in vitro*. *Proc. Nat. Acad. Sci. USA* **88**: 11 295- 11 299

Roemer, T., Delaney, S. and Bussey, H. (1993) *SKN1* and *KRE6* define a pair of functional homologs encoding putative membrane proteins involved in p-glucan synthesis, *Mol. Cell. Biol.* **13**: 4039-4048

Rogers, M.B., Gilson, P.R., Su, V., McFadden, G.I. and Keeling, P.J. (2006) The complete chloroplast genome of the chlorarachniophyte *Bigeloviella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Mol. Biol. Evol.* **24**: 54–62

Ruiz-Herrera J., Gonzalez-Prieto J.M., Ruiz-Medrano, R. (2002) Evolution and phylogenetic relationships of chitin synthases from yeasts and fungi. *FEMS Yeast Research* **1**: 247-256

Sakagami, H., Kikuchi, K., Takeda, M., Sato, T., Ichikawa, S., Fujimaki, M., Wada, C. and Komatsu, N. (1991) Macrophage stimulation activity of antimicrobial N,N-dimethylaminoethyl paramylon. *In Vivo* **5**: 101-106

Sakagami, H., Unten, S., Hanaoka, A., Ohsawa, N., Fujimaki, M., Komatsu, N., Ichikawa, S., Hata, N., Kondo, H. and Konno, K. (1989) Chemical modification potentiates paramylon induction of antimicrobial activity. *In Vivo* **3**: 243-248

Sambrooke, J and Russel, T. (2001) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, New York 14.17- 14.18

Sambrooke, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: A laboratory Manual. Cold Harbor Laboratory Press, Cold Spring Harbor, New York

Samuels, A.L., Giddings, T.H. and Staehelin, L.A. (1995) Cytokinesis in tobacco BY-2 and root tip cells: a new model of cell plate formation in higher plants. *J. Cell Biol.* **130**: 1345–57

Saxena, I. M. and Brown, R. M. Jr. (1995) Identification of a second cellulose synthase gene (*acsAII*) in *Acetobacter xylinum*. *J. Bacteriol.* **177**: 5276–5283

Saxena, I.M. and Brown, R.M. (1997) Identification of cellulose synthase(s) in higher plants: sequence analysis of processive β -glycosyltransferases with the common motif “D, D, D35Q(RQ)XRW”, *Cellulose* **4**: 33–49.

Schols, D., Baba, M., Pauwels, R. and De Clercq, E. (1989) Flow cytometric method to demonstrate whether anti-HIV-I agents inhibit virion binding to T4 + cells. *J. Acquit. Immun. Defic. Syndr.* **2**: 10-15

Seto, N.O., Compston, C.A., Evans, S.V., Bundle, D.R., Narang, S.A. and Palcic, M.M. (1999). Donor substrate specificity of recombinant human blood group A, B and hybrid A/B glycosyltransferases expressed in *Escherichia coli*. *Eur. J. Biochem.* **259**: 770–775

Shematek, E.M. and Cabib, E. (1980) Biosynthesis of the yeast cell wall. *J. Biol. Chem.* **255**: 895-902

Shin, H. and Brown, R.M. Jr. (1998) The first biochemical identification of yeast β -1,3-glucan synthase homolog in cotton fibers (*Gossypium hirsutum*). *Proc. Keystone Symp. Plant Cell Biol., Taos, NM*, p. 45.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal Chem*, **68**:850-858)

Simpson, A.G.B. (1997) The identity and composition of the Euglenozoa. *Arch Protistenkd* **148**: 318– 328

Simpson, A.G.B, van den Hoff, J., Bernhard, C., Burton, H.R. and Patterson, D.J. (1997) The ultrastructure and systematic position of the euglenozoon *Postgaardi mariagerensis*, Fenchel *et al.* *Arch Protistenkd* **147**: 213–225

- Smith, K.A., Lachman, L.B., Oppenheim, J.J. and Favata, M. (1980) The functional relationship of interleukin 2-receptor expression by thymoma cells, *J. Immunol.* **137**: 1226-1231
- Staehelin, L.A. and Giddings, T.H. (1982) Membrane-mediated control of cell wall microfibrillar order. In Subtelny, S. and Green, P. B. [Eds.] *Developmental Order: Its Origin and Regulation*. Alan R. Liss, New York, pp. 133-147
- Stasinopoulos, S.J., Fisher, P.R., Stone, B.A. and Stanisich, V.A. (1999) Detection of two loci involved in (1,3)-b-glucan (curdlan) biosynthesis by *Agrobacterium* sp. ATCC31749, and comparative sequence analysis of the putative curdlan synthase gene. *Glycobiology* **9**: 31–41
- Stoebe, B. and Maier, U.G. (2002) One, two, three: nature's toolbox for building plastids. *Protoplasma* **219**: 123–130
- Sugiyama, A., Suzuki, K., Mitra, S., Arashida, R., Yoshida, E., Nakano, R., Yabuta, Y., Takeuchi, T. (2009) Hepatoprotective effects of paramylon, a beta-1,3-D-glucan isolated from *Euglena gracilis* Z, on acute liver injury induced by carbon tetrachloride in rats. *J. Vet. Med. Sci.*, **7**: 885-890
- Sutherland, I. W. (1998) Novel and established applications of microbial polysaccharides. *Trends in Biotechnology* **16**: 41–46
- Takeyama, H., Kanamaru, A., Yoshino, Y., Kakuta, H., Kawamura, Y. and Matsunaga, T. (1997) Production of antioxidant vitamins, β -carotene, vitamin C, vitamin E, by two-step culture of *Euglena gracilis* Z. *Biotech. Bioengng* **53**: 185–190
- Tarbouriech, N., Charnock, S.J. and Davies, G.J. (2001) Three-dimensional structures of the Mn and Mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: a comparison with related NDPsugar glycosyltransferases. *J. Mol. Biol.* **314**: 655–661
- Thelen, M.P. and Delmer, D.P. (1986) Gel-electrophoretic separation, detection, and characterization of plant and bacterial UDP-glucose glucosyltransferases. *Plant Physiol.* **81**: 913–918
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research.* **22**: 4673- 4680
- Timmis, J.N., Ayliffe, M.A., Huang, C.Y. and Martin, W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**: 123–135
- Tomos, A.D. and Northcote, D.H. (1978) A protein-glucan intermediate during paramylon synthesis. *Biochem. J.* **174**: 283–290

Trepel, F. (2004) Dietary fiber: More than matter of dietetics. II Application in prevention and therapy. Karlsruhe, Germany. *Wiener Klinische Wochenschrift* **116**: 511–522

Turner, A., Basic, A., Harris, P.J. and Reed, S.M. (1998) Membrane fractionation and enrichment of callose synthase from pollen tubes of *Nicotiana glauca* Link et Otto. *Planta* **205**: 380–388

Ueno, R. and Kuno, S. (1987) Dextran sulphate, a potent anti-HIV agent in vitro having synergism with zidovudine. *Lancet* **1**: 1379

Ünlügil, U., Zhou, S., Yuwaraj, S., Sarkar, M., Schachter, H. and Rini, J. (2000) X-ray crystal structure of rabbit N-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily. *EMBO J.* **19**: 5269–5280

Ujita, M., Katsuno, Y., Suzuki K., Sugiyama, K., Takeda, E., Yokoyama, E. and Hara, A. (2006) Molecular cloning and sequence analysis of the β -1,3-glucan synthase catalytic subunit gene from a medicinal fungus, *Cordyceps militaris*. *Mycoscience*, **47**: 98–105

Valdivieso, M.H., Mol, P.C., Shaw, J.A., Cabib, E. and Durh, A. (1991) CAL1, a gene required for activity of chitin synthase 3 in *Saccharomyces cerevisiae*. *J. Cell. Biol.* **144**: 101-109

van den Hoek, C., Mann, D.G. and Jahns, H.M. (1995) Algae. In introduction to phylogeny. Cambridge University Press, Cambridge, UK.

van der Merwe, L. (2007).UDP-glucose: β -(1-3)-glucan (paramylon) synthase form *Euglena gracilis*. Unpublished master's thesis, University of Stellenbosch, Stellenbosch, South Africa

Walne, P.L. and Arnott, H.J. (1967). The comparative ultrastructure and possible function of eyespots: *Euglena granulata* and *Chlamydomonas eugametos*. *Planta* **77**: 325-353

Walne, P.L. and Kivic, P.A. (1990) *Euglenida*. In: Margulis, L., J.O. Corliss, M. Melkonian, and D.J. Chapman, eds. 1990. Handbook of the Protoctista; the structure, cultivation, habits and life histories of the eukaryotic microorganisms and their descendants exclusive of animals, plants and fungi. Jones and Bartlett Publishers. Boston. pp. 270-287

Waterkeyn, L. (1981) Cytochemical localization and function of the 3-linked clucan callose in the developing cotton fibre cell wall. *Protoplasma* **106**: 49-67

Witvrouw, M., Schols, D., Andrei, G., Snoeck, R., Hosoya, M., Pauwels, R., Balzarini, J. and De Clercq, E. (1991) Antiviral activity of low-MW dextran sulphate (derived from dextran MW 1000) compared to dextran sulphate samples of higher MW. *Antiviral Chem. Chemother.* **2**: 171-179

Wong, G.G., Witek-Gianotti, J.S., Temple, P.A., Kriz, R., Ferenz, C., Hewick, R.M., Clark, S.C., Ikebuchi, K., Ogawa, M. (1988) Stimulation of murine hemopoietic colony formation by human IL-6, *J. Immunol.* **140**: 3040-3044

Wrabl, J.O. and Grishin, N.V. (2001) Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily. *J. Mol. Biol.* **314**: 365–374

Yamane, Y., Utsunomiya, T., Watanabe, M. and Sasaki, K. (2001) Biomass production in mixotrophic culture of *Euglena gracilis* under acidic condition and its growth promoting energetic. *Biotechnology Letters* **23**: 1223-1228

Zvelebil, M.J.J.M. and Sternberg, M.J.E. (1988) Analysis and prediction of the location of catalytic residues in enzymes. *Protein Eng.* **2**: 127–138

Addendum A: Predicted peptide LCMSMS sequences of the purified paramylon synthase proteins.

Peptide sequence	NCBI Reference	E-value
95 kDa		
SLDSSGMK	ref YP_001866300.1 putative glycosyl transferase [<i>Nostoc punctiforme</i> PCC 73102]	3018
APIALSPR	ref YP_001557624.1 glycosyl transferase family protein [<i>Clostridium phytofermentans</i> ISDg]	1676
	ref XP_001915260.1 PREDICTED: glycosyltransferase-like 1B [<i>Equus caballus</i>]	5434
DYAAGATR	ref ZP_05280251.1 glycosyl transferase [<i>Bacteroides fragilis</i> 3_1_12]	9784
FAGKAAFL	ref YP_755775.1 glycosyl transferase family protein [<i>Maricaulis maris</i> MCS10]	1676
YDAANDR	ref YP_002785459.1 putative peptidoglycan glycosyltransferase, precursor [<i>Deinococcus deserti</i>]	3544
APIAALIR	ref NP_681033.1 putative glycosyltransferase [<i>Thermosynechococcus elongatus</i> BP-1]	1676
APLAALLR	ref YP_191427.1 putative glycosyltransferase [<i>Gluconobacter oxydans</i> 621H]	287
	ref ZP_04606480.1 glycosyl transferase family 2 [<i>Micromonospora</i> sp. ATCC 39149]	1676
VATVSLPR	ref ZP_05059660.1 glycosyl transferase, group 2 family [<i>Verrucomicrobiae bacterium</i> DG1235]	7291
AVLSSLPR	ref ZP_01879731.1 glucosyltransferase MdoH [<i>Roseovarius</i> sp. TM1035]	1676
GLEAALLR	ref ZP_01302004.1 glucosyltransferase MdoH [<i>Sphingomonas</i> sp. SKA58]	1676
GSLCPPER	ref YP_002951700.1 putative glycosyltransferase [<i>Desulfovibrio magneticus</i> RS-1]	2249
	ref NP 579090.1 glycosyl transferase [<i>Pyrococcus furiosus</i> DSM 3638]	2249
LDEQLRSR	ref YP_003120742.1 glycosyl transferase group 1 [<i>Chitinophaga pinensis</i> DSM 2588]	517
	emb CAD97418.1 putative transmembrane glycosyltransferase [<i>Branchiostoma floridae</i>]	517
EAADDDIVQALNQYTK	ref YP_001684350.1 glycosyl transferase family protein [<i>Caulobacter</i> sp. K31]	146
DLAKQSFIPSDIEENNFQK	ref YP_003293302.1 putative glycosyltransferase [<i>Lactobacillus johnsonii</i> FI9785]	80
80kDa		
APIALSPR	ref YP_001557624.1 glycosyl transferase family protein [<i>Clostridium phytofermentans</i> ISDg]	1678
	ref XP_001915260.1 PREDICTED: glycosyltransferase-like 1B [<i>Equus caballus</i>]	5438
APIAALIR	ref NP_681033.1 putative glycosyltransferase [<i>Thermosynechococcus elongatus</i> BP-1]	1678
APLAALLR	ref YP_191427.1 putative glycosyltransferase [<i>Gluconobacter oxydans</i> 621H]	287
	ref ZP_04606480.1 glycosyl transferase family 2 [<i>Micromonospora</i> sp. ATCC 39149]	1678
GSHAIIR	ref XP_002523261.1 UDP-glucosyltransferase, putative [<i>Ricinus communis</i>]	7296
LAPAALLR	ref YP_001221275.1 putative glycosyl transferase [<i>Clavibacter michiganensis subsp. michiganensis</i>]	2251
LPAALALR	ref YP_003265416.1 glycosyl transferase family 2 [<i>Haliangium ochraceum</i> DSM 14365]	2251
LADALSPR	ref YP_002480849.1 glycosyl transferase group 1 [<i>Desulfovibrio desulfuricans</i> ATCC 27774]	1250
VATVSLPR	ref ZP_05059660.1 glycosyl transferase, group 2 family protein [<i>Verrucomicrobiae bacterium</i> DG1235]	7291
VAERDPR	ref YP_003182395.1 glycosyl transferase family 2 [<i>Eggerthella lenta</i> DSM 2243]	3544
	ref ZP_04055146.1 glycosyl transferase, family 2 [<i>Porphyromonas uenonis</i> 60-3]	3544
	ref YP_001877371.1 glycosyl transferase family 2 [<i>Akkermansia muciniphila</i> ATCC BAA-835]	3544
	ref YP_001611499.1 glycosyltransferase [<i>Sorangium cellulosum</i> 'So ce 56']	3544
VEAAISPR	ref YP_001241524.1 putative glycosyl transferase, group 1 [<i>Bradyrhizobium</i> sp. BTai1]	1676

EAVAAEPR	ref ZP_05803199.1 glycosyl transferase family 2 [<i>Streptomyces flavogriseus</i> ATCC 3331]	1676
GLEAALLR	ref ZP_05782651.1 putative glycosyl transferase [<i>Citricella</i> sp. SE45]	1678
ISAVLPSR	ref YP_645053.1 glycosyl transferase family protein [<i>Rubrobacter xylanophilus</i> DSM 9941]	1249
ISAVILVK	ref NP_908016.1 glycosyltransferase [<i>Wolinella succinogenes</i> DSM 1740]	214
	ref ZP_04632793.1 Glycosyl transferase, group 2 family protein [<i>Yersinia frederiksenii</i> ATCC 33641]	2249
	ref YP_566838.1 polysaccharide biosynthesis protein [<i>Methanococcoides burtonii</i> DSM 6242]	2249
	ref XP_002379617.1 chitin synthase, putative [<i>Aspergillus flavus</i> NRRL3357]	3018
	ref ZP_04808199.1 glycosyltransferase [<i>Helicobacter pullorum</i> MIT 98-5489]	4050
AEAVVVVR	ref YP_002565735.1 polysaccharide biosynthesis protein [<i>Halorubrum lacusprofundi</i> ATCC 49239]	1250
EAAVLALR	ref NP_001154712.1 ATGSL12; 1,3-beta-glucan synthase/transferase [<i>Arabidopsis thaliana</i>]	2251
SLAVEPAR	emb CAI30069.1 glycosyltransferase [<i>Triticum aestivum</i>]	1250
LKDGPNMAK	ref ZP_05625831.1 glycosyl transferase, group 1 [<i>Campylobacter gracilis</i> RM3268]	2532
	ref ZP_04947215.1 Glycosyltransferase [<i>Burkholderia dolosa</i> AUO158]	2532
	ref ZP_02952387.1 capsular polysaccharide biosynthesis protein [<i>Clostridium perfringens</i> D str.]	4559
ELSVNINGK	ref ZP_05390096.1 glycosyl transferase family 2 [<i>Clostridium carboxidivorans</i> P7]	582
IEWNGNIK	ref ZP_01550917.1 putative bifunctional glycosyltransferase, [<i>Stappia aggregata</i> IAM 12614]	932
TLLTLTSE	ref YP_003327119.1 glycosyl transferase family 2 [<i>Xylanimonas cellulosilytica</i> DSM 15894]	1406
TIMIKSNVK	ref YP_695362.1 glycosyl transferase, group 2 family protein [<i>Clostridium perfringens</i> ATCC 13124]	1406
	ref NP_896548.1 glycosyltransferase [<i>Synechococcus</i> sp. WH 8102]	1406
	gb ABV65023.1 glycosyltransferase [<i>Dorosoma cepedianum</i>]	2532
DVLVHPPTR	gb EEQ85757.1 UDP-glucose:sterol glycosyltransferase [<i>Ajellomyces dermatitidis</i> ER-3]	180
	ref XP_002629213.1 UDP-glucose:sterol glycosyltransferase [<i>Ajellomyces dermatitidis</i> SLH14081]	180
TLLDAFLIK	ref YP_001839410.1 putative glycosyltransferase [<i>Leptospira biflexa</i> serovar Patoc strain 'Patoc 1]	582
DKEGEEGNR	dbj BAF75879.1 glucosyltransferase [<i>Dianthus caryophyllus</i>]	2532
HLELPAPLR	ref ZP_02363972.1 glycosyl transferase, group 1 family protein, putative [<i>Burkholderia oklahomensis</i>]	781
	ref ZP_02926586.1 Glycosyl transferase, group 1 family protein [<i>Verrucomicrobium spinosum</i> DSM]	1048
	ref NP_868749.1 polysaccharide biosynthesis protein [<i>Rhodopirellula baltica</i> SH 1]	1048
LDVYVAPRI	ref ZP_03921336.1 possible group 1 glycosyl transferase [<i>Corynebacterium pseudogenitalium</i> ATCC]	582
	ref YP_002760198.1 putative glycosyltransferase [<i>Gemmatimonas aurantiaca</i> T-27]	582
MRPAVDGGSR	ref YP_003112418.1 Peptidoglycan glycosyltransferase [<i>Catenulispora acidiphila</i> DSM 44928]	2831
DVNGGGSPEK	ref YP_446025.1 cellulose-binding domain-containing protein [<i>Salinibacter ruber</i> DSM 13855]	2091
RTDGSFDLIR	ref YP_003143811.1 glycosyl transferase [<i>Slackia heliotrinireducens</i> DSM 20476]	868
AILDQVKVYTK	ref ZP_06061385.1 glycosyltransferase [<i>Streptococcus</i> sp. 2_1_36FAA]	267
SCWDNLSTNSR	ref YP_001551180.1 cell wall biosynthesis glycosyltransferase [<i>Prochlorococcus marinus</i> str. MIT 9211]	645
HTSSFLHPSNR	ref NP_193283.2 UGT84A1; UDP-glycosyltransferase/ sinapate 1-glucosyltransferase [<i>A. thaliana</i>]	645
AEDWKAQVAR	ref YP_002959931.1 Glycosyltransferase [<i>Thermococcus gammatolerans</i> EJ3]	480
MALITTISAPER	ref YP_003179855.1 glycosyl transferase family 2 [<i>Atopobium parvulum</i> DSM 20469]	1553
EADIDRTLTR	ref ZP_05077132.1 glycosyl transferase, group 2 family protein [<i>Rhodobacterales bacterium</i> Y4I]	267
MLIPVMRWQR	ref YP_001634862.1 glycosyl transferase group 1 [<i>Chloroflexus aurantiacus</i> J-10-fl]	482
VAGARFCAHIER	ref YP_114970.1 glycosyl transferase, group 1 family protein [<i>Methylococcus capsulatus</i> str. Bath]	863
RAGAVVDFIPER	ref YP_001234644.1 glycosyl transferase WecB/TagA/CpsF [<i>Acidiphilium cryptum</i> JF-5]	198
NINLEDEKINK	ref ZP_04822410.1 glycosyltransferase [<i>Clostridium botulinum</i> E1 str.]	199
VQARFDQSHNK	ref XP_001944335.1 PREDICTED: similar to glucosyl/glucuronosyl transferases [<i>Acyrtosiphon pisum</i>]	2805
ALMENGLHEMR	ref XP_002301856.1 cellulose synthase [<i>Populus trichocarpa</i>]	198
	gb AAN28294.1 cellulose synthase 2 [<i>Gossypioideis kirkii</i>]	198

	gb AAN28291.1 cellulose synthase 2 [<i>Gossypium raimondii</i>]	198
	gb AAN28290.1 cellulose synthase 2 [<i>Gossypium herbaceum</i>]	198
	gb AAN28293.1 cellulose synthase 2 [<i>Gossypium barbadense</i>]	198
	gb AAO15532.1 AF458083_1 cellulose synthase [<i>Arabidopsis thaliana</i>]	198
	gb AAB37767.1 cellulose synthase [<i>Gossypium hirsutum</i>]	198
	dbj BAB09063.1 cellulose synthase catalytic subunit-like protein [<i>Arabidopsis thaliana</i>]	198
	ref NP_199216.2 CESA4; cellulose synthase/ transferase, [<i>Arabidopsis thaliana</i>]	198
ALYPEVAVAVVAR	ref YP_001381596.1 glycosyl transferase group 1 [<i>Anaeromyxobacter</i> sp. Fw109-5]	110
	ref YP_383895.1 peptidoglycan glycosyltransferase [<i>Geobacter metallireducens</i> GS-15]	265
AGAAALTEKSWPR	ref YP_001819422.1 glycosyl transferase group 1 [<i>Opitutus terrae</i> PB90-1]	147
	ref YP_319626.1 glycosyl transferase family protein [<i>Nitrobacter winogradskyi</i> Nb-255]	198
	ref YP_002499497.1 glycosyl transferase group 1 [<i>Methylobacterium nodulans</i> ORS 2060]	478
QKTPAFDYFIK	ref ZP_04778607.1 family 2 glycosyltransferase [<i>Sphingobacterium spiritivorum</i> ATCC 33861]	1161
	ref YP_957580.1 glycosyl transferase family protein [<i>Marinobacter aquaeolei</i> VT8]	481
	ref YP_001341511.1 glycosyl transferase family protein [<i>Marinomonas</i> sp. MWYL1]	481
QLRGVISSLEQK	ref YP_669975.1 putative glycosyltransferase [<i>Escherichia coli</i> 536]	863
	ref ZP_03031752.1 mannosyl transferase [<i>Escherichia coli</i> F11]	863
LASAMGAPGLAGDAR	ref ZP_03450438.1 glycosyl transferase, group 2 family protein [<i>Burkholderia pseudomallei</i> 576]	264
	ref ZP_04889698.1 glycosyl transferase, group 2 family protein [<i>Burkholderia pseudomallei</i> 1655]	264
	ref ZP_02509252.1 putative cellulose synthase [<i>Burkholderia pseudomallei</i> BCC215]	264
	ref YP_337461.1 beta 1,3 glucan synthase catalytic subunit [<i>Burkholderia pseudomallei</i> 1710b]	264
TGREGTTGYSSNK	gb AAN01252.1 Unknown protein similar to putative cellulose synthase [<i>Oryza sativa</i>]	0.007
	gb ABB47240.2 Cellulose synthase family protein, expressed [<i>Oryza sativa</i> (japonica cultivar-group)]	0.007
	ref XP_567874.1 beta-1,3 glucan biosynthesis-related protein [<i>Cryptococcus neoformans</i>]	2077
YEFLQNRECMK	ref ZP_06086789.1 glycosyltransferase family 2 [<i>Bacteroides</i> sp. 3_1_33FAA]	358
HASTGDLTTIQDTK	ref YP_002430018.1 glycosyl transferase family 2 [<i>Desulfatibacillum alkenivorans</i> AK-01]	858
	ref ZP_05922740.1 glycosyl transferase, family 2 [<i>Enterococcus faecium</i> TC 6]	1151

72 kDa

LLGSRGFNAVSR	ref YP_168098.1 glucosyltransferase MdoH [<i>Ruegeria pomeroyi</i> DSS-3]	1562
	gb ABC25314.1 glucans biosynthesis glucosyltransferase H [uncultured marine bacterium Ant24C4]	2096
	ref ZP_05122270.1 glucans biosynthesis glucosyltransferase H [<i>Rhodobacteraceae bacterium</i> KLH11]	0.071
	ref ZP_01055929.1 glucosyltransferase MdoH [<i>Roseobacter</i> sp. MED193]	268
	ref YP_002550848.1 glucosyltransferase [<i>Agrobacterium vitis</i> S4]	868
	ref ZP_05787035.1 glucans biosynthesis glucosyltransferase H [<i>Silicibacter lacuscaerulensis</i> ITI-1157]	1164
	ref YP_001262990.1 polysaccharide biosynthesis protein [<i>Sphingomonas wittichii</i> RW1]	1164

62 kDa

MALITTISAPER	ref YP_003179855.1 glycosyl transferase family 2 [<i>Atopobium parvulum</i> DSM 20469]	1154
ISMIEGGLVDLR	ref ZP_05594618.1 polysaccharide biosynthesis protein [<i>Enterococcus faecalis</i> AR01/DG]	266
	ref ZP_03950093.1 PST family polysaccharide transporter [<i>Enterococcus faecalis</i> TX0104]	266

IVCDLKVGDNVK	ref YP_001353521.1 glucan biosynthesis protein D [<i>Janthinobacterium</i> sp. Marseille]	357
	ref YP_001918677.1 nucleotide sugar dehydrogenase [<i>Natranaerobius thermophilus</i> JW/NM-WN-LF]	863
LSVEEMRGLIR	ref ZP_02931446.1 glycosyltransferase [<i>Verrucomicrobium spinosum</i> DSM 4136]	267
AQASVSQTQNLK	ref NP_200715.2 glycosyl transferase family 1 protein [<i>Arabidopsis thaliana</i>]	863
EENVNVHDVGNNTIK	ref YP_003307971.1 Peptidoglycan glycosyltransferase [<i>Sebaldella termitidis</i> ATCC 33386]	109
MNVGIVVHGPEIIDSGFAEK	ref YP_002938628.1 Glycosyltransferase Family 2 modular protein [<i>Eubacterium rectale</i> ATCC 33656]	194
QYGSVANYLEKGIGITAAEK	ref ZP_06026371.1 glycosyl transferase group 1 [<i>Fusobacterium periodonticum</i> ATCC 33693]	144
ISSHSLEELAAELQISKEK	ref YP_001450314.1 glycosyltransferase [<i>Streptococcus gordonii</i> str. Challis substr. CH1]	14

60 kDa		
-		
ICSEGAKIL	ref XP_001845522.1 glucosyl/glucuronosyl transferase [<i>Culex quinquefasciatus</i>]	1409
	ref XP_001663166.1 glucosyl/glucuronosyl transferases [<i>Aedes aegypti</i>]	2537
MIKQQSAK	ref XP_001827919.1 nucleotide-sugar transporter [<i>Enterocytozoon bieneusi</i> H348]	2255
QLGDYIPK	ref ZP_03131719.1 glycosyl transferase group 1 [<i>Chthoniobacter flavus</i> Ellin428]	3026
MIADVVSAT	ref YP_949787.1 putative glycosyl transferase, group 1 family protein [<i>Arthrobacter aurescens</i> TC1]	2537
FTRDWTG	ref YP_001404698.1 glycosyl transferase, group 1 [<i>Candidatus Methanoregula boonei</i> 6A8]	1970
FTRSGTER	ref YP_002753948.1 glycosyl transferase, group 2 family [<i>Acidobacterium capsulatum</i> ATCC 51196]	1250
TFRMQDR	ref YP_460432.1 glycosyltransferase [<i>Syntrophus aciditrophicus</i> SB]	2643
KIAHSLER	dbj BAB86933.1 glycosyltransferase like protein [<i>Vigna angularis</i>]	932
AAAATGAAGHR	ref YP_002230518.1 cellulose synthase regulator protein [<i>Burkholderia cenocepacia</i> J2315]	1561
	emb CAR51686.1 cyclic di-GMP binding protein [<i>Burkholderia cenocepacia</i> J2315]	1561
LQAAARPAR	ref ZP_02204800.1 glycosyl transferase, group 1 [<i>Dehalococcoides</i> sp. VS]	324
	ref YP_336071.1 putative glycosyltransferase [<i>Burkholderia pseudomallei</i> 1710b]	1891
LQAAIPVK	ref YP_001279140.1 glycosyl transferase, group 1 [<i>Psychrobacter</i> sp. PRwf-1]	1050
LSLLSQKDSLK	ref ZP_03560622.1 polysaccharide biosynthesis protein [<i>Glaciecola</i> sp. HTCC2999]	265
AMMDLAVPVLQADLK	ref YP_001748263.1 glycosyl transferase group 1 [<i>Pseudomonas putida</i> W619]	147
DSILLYVTRMLK	gb ABP96910.1 cellulose synthase 1 [<i>Phytophthora ramorum</i>]	147
QVNSIYGKFSASWK	ref YP_003182856.1 glycosyl transferase family 51 [<i>Eggerthella lenta</i> DSM 2243]	1151
SDLDAALSELSVALGEVQR	ref YP_002128859.1 glycosyltransferase [<i>Phenylobacterium zucineum</i> HLK1]	10
GVCIKTINQGLTFSYSNK	ref ZP_05493841.1 glycosyltransferase 36 [<i>Clostridium papyrosolvens</i> DSM 2782]	472
CAAVPTTSDFHTEMAQYVFTMNRVGK	ref YP_622459.1 putative ABC transporter ATP-binding protein [<i>Burkholderia cenocepacia</i> AU 1054]	4.00E-17
RNSGWVFENPSIGVLELWVLATNFR	ref XP_001930359.1 glucan 1,3-beta-glucosidase [<i>Pyrenophora tritici-repentis</i> Pt-1C-BFP]	44
KAADVAVVYADAAGDLNTVFDNLPTINK	ref ZP_05649292.1 glycosyltransferase [<i>Enterococcus gallinarum</i> EG2]	7.5
	ref NP_948694.1 putative exopolysaccharide polymerization protein [<i>Rhodopseudomonas palustris</i>]	44

50 kDa		
CMSSHI	ref ZP_02926473.1 glycosyl transferase, group 1 family protein [<i>Verrucomicrobium spinosum</i>]	9905
EHA AHL	ref YP_002129322.1 glycosyl transferase, group 2 family protein [<i>Phenylobacterium zucineum</i> HLK1]	4100
MHCMR	ref ZP_05476892.1 glycosyltransferase probably involved in cell wall biogenesis [<i>Streptomyces</i> sp. AA4]	11076
	emb CBH17274.1 endosomal integral membrane protein, putative [<i>Trypanosoma brucei gambiense</i>]	11076
MHDFK	ref ZP_03967889.1 possible glycosyltransferase [<i>Sphingobacterium spiritivorum</i> ATCC 33300]	4585

MHFDK	ref XP_002174225.1 cell wall alpha-1,3-glucan synthase mok12 [<i>Schizosaccharomyces japonicus</i> yFS275]	4585
GHGGGRQR	ref YP_003324970.1 glycosyl transferase group 1 [<i>Xylanimonas cellulosilytica</i> DSM 15894]	9836
HNGGVGRK	gb EEH54761.1 glycosyltransferase family 1 protein [<i>Micromonas pusilla</i> CCMP1545]	13198
GHGGGGARR	ref NP_001151784.1 LOC100285419 [<i>Zea mays</i>] ; gb ACG44302.1 glycosyl transferase	785
GPPGDIIR	ref ZP_01463583.1 glycosyl transferase, group 1 family protein [<i>Stigmatella aurantiaca</i> DW4/3-1]	2262
GSHAIAIR	ref XP_002523261.1 UDP-glucosyltransferase, putative [<i>Ricinus communis</i>]	7336
IGEAALLR	ref ZP_04659432.1 family 2 glycosyl transferase [<i>Selenomonas flueggei</i> ATCC 43531]	2255
	ref ZP_00047747.2 COG2943: Membrane glycosyltransferase [<i>Magnetospirillum magnetotacticum</i>]	2255
VATVSLPR	ref ZP_05059660.1 glycosyl transferase, group 2 family protein [<i>Verrucomicrobiae bacterium</i> DG1235]	7311
AVLSSLPR	ref ZP_01879731.1 glucosyltransferase MdoH [<i>Roseovarius</i> sp. TM1035]	1681
IGEAAPER	ref ZP_03129620.1 glycosyl transferase group 1 [<i>Chthoniobacter flavus</i> Ellin428]	4061
EAVAAEPR	ref ZP_05803199.1 glycosyl transferase family 2 [<i>Streptomyces flavogriseus</i> ATCC 33331]	1685
	ref YP_003290600.1 glycosyl transferase group 1 [<i>Rhodothermus marinus</i> DSM 4252]	1685
IGEAAPER	ref ZP_03129620.1 glycosyl transferase group 1 [<i>Chthoniobacter flavus</i> Ellin428]	4072
ISLGAEPR	ref YP_001236071.1 glycosyl transferase, group 1 [<i>Acidiphilium cryptum</i> JF-5]	5463
IGEAALLR	ref ZP_04659432.1 family 2 glycosyl transferase [<i>Selenomonas flueggei</i> ATCC 43531]	2257
	ref ZP_00047747.2 COG2943: Membrane glycosyltransferase [<i>Magnetospirillum magnetotacticum</i>]	2257
AEGIAEPR	ref ZP_01460876.1 glycosyl transferase, group 2 family protein [<i>Stigmatella aurantiaca</i> DW4/3-1]	5463
XLVASIPR	gb EEU48017.1 glycosyltransferase family 4 [<i>Nectria haematococca</i> mpVI 77-13-4]	5463
ELGAAEPR	ref ZP_01048515.1 probable glycosyl transferase [<i>Nitrobacter</i> sp. Nb-311A]	5463
SLGLAEPR	ref XP_001691334.1 glycosyltransferase-like protein [<i>Chlamydomonas reinhardtii</i>]	2262
AEAVVVVR	ref YP_002565735.1 polysaccharide biosynthesis protein [<i>Halorubrum lacusprofundi</i> ATCC 49239]	1257
ASLVVVVR	gb ABF99599.1 glycosyl transferase, group 2 family protein, expressed [<i>Oryza sativa</i>]	3037
DRVVPTR	ref ZP_03971206.1 glycosyltransferase [<i>Corynebacterium glucuronolyticum</i> ATCC 51866]	4784
DRVVVVR	ref XP_001325769.1 glycosyl transferase [<i>Trichomonas vaginalis</i> G3]	2657
IGEAAPER	ref ZP_03129620.1 glycosyl transferase group 1 [<i>Chthoniobacter flavus</i> Ellin428]	4075
KSEIIGPV	ref YP_002362773.1 glycosyl transferase group 1 [<i>Methylocella silvestris</i> BL2]	5467
VATVSLPR	ref ZP_05059660.1 glycosyl transferase, group 2 family protein [<i>Verrucomicrobiae bacterium</i> DG1235]	7311
VAISALIR	ref YP_002950715.1 polysaccharide biosynthesis protein [<i>Geobacillus</i> sp. WCH70]	2255
VAISLPSR	ref YP_002158555.1 cellulose synthase operon protein YhjU [<i>Vibrio fischeri</i> MJ11]	2255
	ref ZP_03154586.1 glycosyl transferase family 9 [<i>Cyanotheca</i> sp. PCC 7822]	5448
VATVSLPR	ref ZP_05059660.1 glycosyl transferase, group 2 family protein [<i>Verrucomicrobiae bacterium</i> DG1235]	7315
QLVLVSPS	ref ZP_03888526.1 PMT family glycosyltransferase [<i>Geodermatophilus obscurus</i>]	1253
VAAEPLSR	ref YP_743172.1 glycosyl transferase family protein [<i>Alkalilimnicola ehrlichii</i> MLHE-1]	4063
VAEALVLK	ref ZP_04054840.1 glycosyl transferase, group 1 family protein [<i>Porphyromonas uenonis</i> 60-3]	2257
DGKPRAAK	ref XP_001939736.1 GPI-anchored cell wall beta-1,3-endoglucanase EglC [<i>Pyrenophora tritici-repentis</i>]	7336
TAQRPNR	ref YP_002469728.1 glycosyl transferase family 2 protein [<i>Bifidobacterium animalis subsp. lactis</i>]	4784
ANYDAILR	ref YP_345320.1 glycosyl transferase family protein [<i>Rhodobacter sphaeroides</i> 2.4.1]	289
LIACADSDK	ref ZP_01902080.1 glycosyltransferase involved in cell wall biogenesis [<i>Roseobacter</i> sp. AzwK-3b]	1054
NATNSVSSR	ref ZP_04611996.1 Glycosyl transferase family 2 [<i>Yersinia rohdei</i> ATCC 43380]	2546
TVSLQAPSSR	gb EEY57179.1 endo-1,3(4)-beta-glucanase, putative [<i>Phytophthora infestans</i> T30-4]	1567
HLELPAPLR	ref NP_868749.1 polysaccharide biosynthesis protein [<i>Rhodopirellula baltica</i> SH 1]	1051
	ref ZP_02363972.1 glycosyl transferase, group 1 family protein, putative [<i>Burkholderia oklahomensis</i>]	783
	ref ZP_02926586.1 Glycosyl transferase, group 1 family protein [<i>Verrucomicrobium spinosum</i> DSM]	1051

YTELPNRLK	ref YP_209926.1 putative glycosyltransferase [<i>Bacteroides fragilis</i> NCTC 9343]	242
AGMDASAIRNK	gb AAN64562.1 glycosyltransferase [<i>Streptococcus gordonii</i>]	870
	ref ZP_04062055.1 glycosyl transferase, group 1 family protein [<i>Streptococcus salivarius</i> SK126]	2102
	ref YP_002951043.1 glycosyl transferase group 1 [<i>Geobacillus</i> sp. WCH70]	2102
AWHLPRILK	ref XP_001300232.1 glycosyl transferase [<i>Trichomonas vaginalis</i> G3]	1054
ALAAALRIQGR	ref ZP_04367828.1 glycosyl transferase [<i>Cellulomonas flavigena</i> DSM 20109]	200
AIKNIINIK	ref ZP_05633447.1 putative glycosyltransferase [<i>Fusobacterium ulcerans</i> ATCC 49185]	62
MGTTAALSQTTI	ref YP_002280725.1 cellulose synthase subunit B [<i>Rhizobium leguminosarum</i> bv. trifolii WSM2304]	2096
GLYKSIVSATR	ref YP_003067580.1 putative glycosyl transferase [<i>Methylobacterium extorquens</i> DM4]	2821
	ref YP_001924069.1 glycosyl transferase group 1 [<i>Methylobacterium populi</i> BJ001]	2821
TYVKLVFSHK	ref ZP_03154914.1 polysaccharide biosynthesis protein [<i>Cyanotheca</i> sp. PCC 7822]	2829
ENENMSNQKK	ref XP_001663167.1 glucosyl/glucuronosyl transferases [<i>Aedes aegypti</i>]	650
TYFMIVER	ref ZP_01127348.1 glucosyltransferase for synthesis of periplasmic glucans [<i>Nitrococcus mobilis</i>]	785
NRCLAELLGLK	ref ZP_04441934.1 glycosyltransferase [<i>Lactobacillus rhamnosus</i> LMS2-1]	1164
	ref ZP_03964155.1 group 1 glycosyl transferase [<i>Lactobacillus paracasei subsp. paracasei</i> ATCC 25302]	1164
	ref YP_001611662.1 glycosyltransferase [<i>Sorangium cellulosum</i> 'So ce 56']	1164
ESEQINKEFTK	gb ACA34492.1 putative glycosyltransferase [<i>Campylobacter jejuni</i>]	650
YLSYSYYCYK	ref ZP_04054743.1 glycosyl transferase, group 2 family protein [<i>Porphyromonas uenonis</i> 60-3]	480
HKGLDTSVVSGPR	ref ZP_01253348.1 glycosyl transferase, group 2 family protein [<i>Psychroflexus torquis</i> ATCC 700755]	1557
NDHAVGEDIRAR	ref ZP_02488140.1 putative glucosyltransferase [<i>Burkholderia pseudomallei</i> NCTC 13177]	83
	ref YP_002894939.1 hopanoid biosynthesis glycosyl transferase HpnI [<i>Burkholderia pseudomallei</i>]	83
	ref ZP_02354145.1 glycosyl transferase, group 2 family protein [<i>Burkholderia oklahomensis</i> EO147]	83
	ref ZP_02462016.1 glycosyl transferase, group 2 family protein [<i>Burkholderia thailandensis</i> MSMB43]	83
	ref YP_104815.1 glycosyl transferase, group 2 family protein [<i>Burkholderia mallei</i> ATCC 23344]	83
AADGPTRFGTGFR	ref YP_003006358.1 cellulose synthase operon C domain protein [<i>Dickeya zeae</i> Ech1591]	483
	ref ZP_05102192.1 glycosyl transferase, group 1 [<i>Roseobacter</i> sp. GAI101]	1161
LTANANTENECKK	gb ACJ63461.1 glucosyltransferase GTF8-2 [<i>Weissella confusa</i>]	645
APSPGPFVPPVSQR	ref YP_925234.1 glycosyl transferase, group 1 [<i>Nocardioides</i> sp. JS614]	357
	ref YP_001994131.1 glycosyl transferase family 2 [<i>Rhodopseudomonas palustris</i> TIE-1]	479
GIHSGTWKLSPPR	ref YP_630159.1 polysaccharide biosynthesis/export protein [<i>Myxococcus xanthus</i> DK 1622]	481
SGFFSHCLSLFSPK	ref ZP_05724441.1 glycosyl transferase group 1 [<i>Dickeya dadantii</i> Ech586]	863
	ref ZP_06115144.1 glycosyl transferase, group 1 [<i>Clostridium hathewayi</i> DSM 13479]	1157
DAVAMFGYLSSDVANK	GENE ID: 374033 UGT8 UDP glycosyltransferase 8 [<i>Gallus gallus</i>]	197
VKGNFVNGLPDGTLEK	ref YP_679052.1 glucosyltransferase-S [<i>Cytophaga hutchinsonii</i> ATCC 33406]	197
	ref ZP_04061534.1 glucosyltransferase-I [<i>Streptococcus salivarius</i> SK126]	147
WDYAVVASGGFFIAGK	ref YP_152483.2 peptidoglycan synthetase [<i>Salmonella enterica subsp. enterica serovar Paratyphi A</i>]	476
VRIVCNSQLQADDVK	ref YP_001126273.1 peptidoglycan binding protein [<i>Geobacillus thermodenitrificans</i> NG80-2]	356
	ref ZP_04557557.1 glycosyl transferase [<i>Bacteroides</i> sp. D4]	478
	ref YP_001735683.1 glycosyl transferase [<i>Synechococcus</i> sp. PCC 7002]	641
	ref YP_283382.1 glycosyl transferase, group 1 [<i>Dechloromonas aromatica</i> RCB]	641
	ref ZP_06060227.1 glycosyl transferase [<i>Streptococcus</i> sp. 2_1_36FAA]	860
LLAGFGNEVLAGYGIGSR	ref YP_003201935.1 UDP-glucuronosyl/UDP-glucosyltransferase [<i>Nakamurella multipartita</i> DSM 44233]	60
GTGCSPDLPPPSRAEMGR	ref YP_001537113.1 glycosyl transferase group 1 [<i>Salinispora arenicola</i> CNS-205]	60
	ref YP_001158948.1 glycosyl transferase, group 1 [<i>Salinispora tropica</i> CNB-440]	60

	ref YP_002481839.1 glycosyl transferase group 1 [<i>Cyanotheca</i> sp. PCC 7425]	196
LITLWQAVVWRVTSR	ref YP_002756046.1 glycosyl transferase, group 2 family [<i>Acidobacterium capsulatum</i> ATCC 51196]	265
AWNDEDPDAFEHR	ref YP_001152946.1 glycosyl transferase family protein [<i>Pyrobaculum arsenaticum</i> DSM 13514]	19
	ref NP_559585.1 glycosyl transferase, putative [<i>Pyrobaculum aerophilum</i> str. IM2]	61
AVGAHHVERTAVFVGDVDR	ref YP_002648411.1 Exopolysaccharide biosynthesis glycosyl transferase WcaL, [<i>Erwinia pyrifoliae</i>]	45
QYPGTMLIVSHDRYLLNK	ref YP_002509151.1 ABC transporter related [<i>Halothermothrix orenii</i> H 168]	3.00E-09

35 kDa

VVPAAAK	ref NP_558604.1 glycosyltransferase (type 1) [<i>Pyrobaculum aerophilum</i> str. IM2]	2700
HGATAAK	gb ACD03254.1 UDP-glycosyltransferase UGT93B9 [<i>Avena strigosa</i>]	21140
QHTRN	ref YP_003269085.1 Glycosyl transferase, family 3-like protein [<i>Haliangium ochraceum</i> DSM 14365]	15100
	ref ZP_05855684.1 glycosyl transferase [<i>Blautia hansenii</i> DSM 20583]	15100
AIDALIAR	ref YP_001749767.1 glycosyl transferase group 1 [<i>Pseudomonas putida</i> W619]	949
ALDALLAR	ref YP_002952721.1 putative glycosyltransferase [<i>Desulfovibrio magneticus</i> RS-1]	293
ANNAILVK	ref XP_001150762.1 UDP glycosyltransferase 1 family, polypeptide A6 isoform 1 [<i>Pan troglodytes</i>]	1274
DALALLAR	ref YP_003156902.1 glycosyl transferase group 1 [<i>Desulfomicrobium baculatum</i> DSM 4028]	1709
VEAAISPR	ref YP_001241524.1 putative glycosyl transferase, group 1 [<i>Bradyrhizobium</i> sp. BTAi1]	1703
	ref YP_841154.1 cell wall biogenesis regulatory protein [<i>Ralstonia eutropha</i> H16]	946
VAKAQAVR	ref YP_002974304.1 glycosyl transferase group 1 [<i>Rhizobium leguminosarum</i> bv. trifolii WSM1325]	3067
VATVSLPR	ref ZP_05059660.1 glycosyl transferase, group 2 family protein [<i>Verrucomicrobiae bacterium</i> DG1235]	7408
VAERDPR	ref YP_003182395.1 glycosyl transferase family 2 [<i>Eggerthella lenta</i> DSM 2243]	3600
	ref ZP_04055146.1 glycosyl transferase, family 2 [<i>Porphyromonas uenonis</i> 60-3]	3600
	ref YP_001877371.1 glycosyl transferase family 2 [<i>Akkermansia muciniphila</i> ATCC BAA-835]	3600
IGEAAPER	ref ZP_03129620.1 glycosyl transferase group 1 [<i>Chthoniobacter flavus</i> Ellin428]	4115
IGEAAPER	ref ZP_03129620.1 glycosyl transferase group 1 [<i>Chthoniobacter flavus</i> Ellin428]	4129
IGEAALLR	ref ZP_04659432.1 family 2 glycosyl transferase [<i>Selenomonas flueggei</i> ATCC 43531]	2294
GLEAALLR	ref ZP_05782651.1 putative glycosyl transferase [<i>Citricella</i> sp. SE45]	1709
	ref ZP_01302004.1 glucosyltransferase MdoH [<i>Sphingomonas</i> sp. SKA58]	1709
AVTVVEPK	ref YP_317158.1 glycosyl transferase family protein [<i>Nitrobacter winogradskyi</i> Nb-255]	163
VAAVTPER	ref ZP_04775004.1 glycosyl transferase family 2 [<i>Allochrocatium vinosum</i> DSM 180]	3077
GGNAAALLR	ref YP_003336149.1 Glycosyltransferase-like protein [<i>Streptosporangium roseum</i> DSM 43021]	2580
EAVAAEPR	ref ZP_05803199.1 glycosyl transferase family 2 [<i>Streptomyces flavogriseus</i> ATCC 33331]	1703
	ref YP_003290600.1 glycosyl transferase group 1 [<i>Rhodothermus marinus</i> DSM 4252]	1703
	ref YP_966185.1 peptidoglycan glycosyltransferase [<i>Desulfovibrio vulgaris</i> subsp.	1703
ELGAAEPR	ref ZP_01048515.1 probable glycosyl transferase [<i>Nitrobacter</i> sp. Nb-311A]	5521
IGEAAPER	ref ZP_03129620.1 glycosyl transferase group 1 [<i>Chthoniobacter flavus</i> Ellin428]	4115
VEAAISPR	ref YP_001241524.1 putative glycosyl transferase, group 1 [<i>Bradyrhizobium</i> sp. BTAi1]	1703
	ref ZP_02371622.1 lipopolysaccharide biosynthesis protein [<i>Burkholderia thailandensis</i>]	705
	ref YP_841154.1 cell wall biogenesis regulatory protein [<i>Ralstonia eutropha</i> H16]	946
LADALSPR	ref YP_002480849.1 glycosyl transferase group 1 [<i>Desulfovibrio desulfuricans</i> subsp. <i>desulfuricans</i> str.]	1269
LDAALSPR	ref ZP_01878060.1 glycosyl transferase, group 1 family protein [<i>Roseovarius</i> sp. TM1035]	1269
	ref YP_110445.1 lipopolysaccharide biosynthesis protein [<i>Burkholderia pseudomallei</i>]	1269
LSSPATLNSR	ref YP_003138348.1 Periplasmic component of Tol biopolymer transport system-like [<i>Cyanotheca</i> sp.]	1570

LEAAADGITGK	ref YP_685025.1 glycosyltransferase (group 1) [uncultured methanogenic archaeon RC-I]	1167
MRLGILGTGK	ref YP_385123.1 glycosyltransferase-like protein [<i>Geobacter metallireducens</i> GS-15]	872
TWLAPDGTGK	gb ABC67269.1 putative glycosyltransferase [<i>Streptomyces hygroscopicus</i> subsp.	1587
EVRIGSASVR	ref YP_003013567.1 Peptidoglycan glycosyltransferase [<i>Paenibacillus</i> sp. JDR-2]	2108
EVVEAAGGSVR	ref YP_003153748.1 predicted glycosyltransferase [<i>Brachybacterium faecium</i> DSM 4810]	200
EAILTQGSVR	ref ZP_04832497.1 glycosyl transferase, group 1 family [<i>Prevotella melaninogenica</i> ATCC 25845]	887
SAVNGEAANIK	ref ZP_05138412.1 glycosyl transferase, group 1 [<i>Prochlorococcus marinus</i> str. MIT 9202]	2866
	ref ZP_02185159.1 glycosyl transferase, group 1 family protein [<i>Carnobacterium</i> sp. AT7]	1186
QKLPIQIVDK	ref YP_684816.1 putative oligosaccharyltransferase [uncultured methanogenic archaeon RC-I]	661
TSKGSAYLNK	ref YP_684816.1 putative oligosaccharyltransferase [uncultured methanogenic archaeon RC-I]	2866
	ref XP_002519441.1 UDP-glucosyltransferase, putative [<i>Ricinus communis</i>]	884
RVSALNIAHLK	gb EES53993.1 putative glycosyl transferase, group 1 [<i>Leptospirillum ferro-diazotrophum</i>]	2136
GPSSSSSADACR	ref ZP_06276423.1 Protein involved in beta-1 3-glucan synthesis-like protein [<i>Streptomyces</i> sp. ACT-1]	488
QVNDQYVSLAK	ref ZP_04493766.1 glycosyltransferase [<i>Spirosoma linguale</i> DSM 74	1592
SFASNTLRGVGR	ref ZP_05402172.1 putative polysaccharide biosynthesis protein [<i>Clostridium difficile</i> QCD-23m63]	1183
DILGTYERVAK	ref NP_227976.1 geranyltranstransferase [<i>Thermotoga maritima</i> MSB8]	1186
	ref YP_001738822.1 polyprenyl synthetase [<i>Thermotoga</i> sp. RQ2]	1186
LHHMSEEILR	ref YP_948600.1 putative glycosyl transferase, group 1 family protein [<i>Arthrobacter aurescens</i> TC1]	493
	ref YP_002514796.1 glycosyl transferase, group 1 [<i>Thioalkalivibrio</i> sp. HL-EbGR7]	887
QHQSSTSENK	ref ZP_01621904.1 Glycosyl transferase, family 2 [<i>Lyngbya</i> sp. PCC 8106]	884
IKATDFMVAMR	ref ZP_03700828.1 glycosyl transferase family 9 [<i>Flavobacteria bacterium</i> MS024-3C]	884
	ref XP_002486496.1 exo-beta-1,3-glucanase, putative [<i>Talaromyces stipitatus</i> ATCC 10500]	884
RTGHVLAASSGK	gb AAH56559.1 Dolichyl-diphosphooligosaccharide-protein glycosyltransferase [<i>Danio rerio</i>]	1179
GLPAGEPEMIIR	ref YP_003122797.1 glycosyl transferase family 2 [<i>Chitinophaga pinensis</i> DSM 2588]	657
	ref YP_003087422.1 Cellulose synthase (UDP-forming) [<i>Dyadobacter fermentans</i> DSM 18053]	657
LQSVPLDELLR	ref YP_001999413.1 glycosyl transferase group 1 [<i>Chlorobaculum parvum</i> NCIB 8327]	491
	ref NP_001063684.1 Os09g0518000 [<i>Oryza sativa</i> (japonica cultivar-group)]	273
GAFHTVSLQKYSK	ref ZP_03395721.1 glycosyl transferase, group 2 family protein [<i>Pseudomonas syringae</i> pv. tomato T1]	1578
DPICIAXSXCCGKR	emb CBH11780.1 ABC transporter, putative [<i>Trypanosoma brucei gambiense</i> DAL972]	362
EYVAIKQVPYQK	GENE ID: 6172670 Exig_1766 glycosyl transferase family protein[<i>Exiguobacterium sibiricum</i> 255-15]	271
QNEYFCFATKSK	emb CAJ70866.1 similar to glycosyl transferase family 1 [<i>Candidatus Kuenenia stuttgartiensis</i>]	879
EDIINRLSNCIR	ref NP_812284.1 glycosyltransferase [<i>Bacteroides thetaiotaomicron</i> VPI-5482]	202
LMILGTPFAAPNWK	ref ZP_05670075.1 glucosyltransferase [<i>Enterococcus faecium</i> 1,231,410]	363
	ref ZP_03374705.1 glucan biosynthesis protein [<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi str.]	487
ANLTASSAKVAHHHD	ref ZP_01063294.1 predicted calcium-binding protein [<i>Vibrio</i> sp. MED222]	485
KGQGSYFNSGTWAR	ref ZP_03074363.1 Peptidoglycan glycosyltransferase [<i>Lactobacillus reuteri</i> 100-23]	363
	ref ZP_04869645.1 glycosyl transferase [<i>Helicobacter canadensis</i> MIT 98-5491]	1176
	gb ABY41241.1 CP7 beta-1,3-1,4-glucanase [<i>Paenibacillus polymyxa</i>]	202
AADADTLQAYTEEVIK	ref XP_001936633.1 polysaccharide synthase [<i>Pyrenophora tritici-repentis</i> Pt-1C-BFP]	360
EYGVYYFSDNPEIAK	ref YP_002544914.1 glycosyltransferase protein [<i>Agrobacterium radiobacter</i> K84]	270
	ref ZP_04061326.1 glucosyltransferase-I [<i>Streptococcus salivarius</i> SK126]	150
RIGSSSDTSSDTSSSSSSSSDK	ref YP_925858.1 glycosyl transferase family protein [<i>Nocardioidea</i> sp. JS614]	69
	ref ZP_03864096.1 glycosyl transferase [<i>Kribbella flavida</i> DSM 17836]	168
	ref ZP_03864798.1 putative glucan synthasis protein [<i>Kribbella flavida</i> DSM 17836]	168

QVYNVIRQLFQIAMPFGSK	ref ZP_05316312.1 glycosyl transferase family 2 [<i>Nitrosomonas</i> sp. AL212]	148
SIAFSVCAAQEALWRISGSSR	ref YP_001537238.1 glycosyl transferase family protein [<i>Salinispora arenicola</i> CNS-205]	265
MPDHLGQPVSQAVAFPNDSK	ref ZP_04775708.1 glycosyl transferase group 1 [<i>Allochrocatium vinosum</i> DSM 180]	265
LQRGSAPYLQFELVVYSR	ref ZP_04773562.1 glycosyl transferase group 1 [<i>Allochrocatium vinosum</i> DSM 180]	25
NREQDHASDAYTALLSSVLQR	ref XP_812757.1 glycosyl transferase [<i>Trypanosoma cruzi</i> strain CL Brener]	145
	ref XP_805961.1 glycosyl transferase [<i>Trypanosoma cruzi</i> strain CL Brener]	145
GLGMLGWVFFPPREDLSGGIDR	ref YP_003129966.1 glycosyl transferase group 1 [<i>Halorhabdus utahensis</i> DSM 12940]	61
	ref ZP_03997417.1 glycosyltransferase [<i>Halogeometricum borinquense</i> DSM 11551]	61
	ref YP_658229.1 LPS glycosyltransferase [<i>Haloquadratum walsbyi</i> DSM 16790]	61
IDSSNSTICEVSPVSLNGENLK	emb CAP39914.1 UDP-GlcNAc diphosphorylase [<i>Trypanosoma brucei brucei</i>]	81
	ref XP_635927.1 UDP-N-acetylglucosamine pyrophosphorylase [<i>Dictyostelium discoideum</i>]	2.00E-12